Theory and Praxis of Map Analysis in CHEF
Part 2: Nonlinear Normal Form

Leo Michelotti

April 10, 2009

1 Introduction

This is the second of three memos describing how normal form map analysis is implemented in CHEF. The first [1] explained the manipulations required to assure that initial, linear transformations preserved Poincaré invariants, thereby confirming correct normalization of action-angle coordinates. In this one, the transformation will be extended to nonlinear terms. The third, describing how the algorithms were implemented within the software of CHEF's libraries, most likely will never be written.

The first section, Section 2, quickly lays out preliminary concepts and relationships. In Section 3, we shall review the perturbation theory - an iterative sequence of transformations that converts a nonlinear mapping into its normal form - and examine the equation which moves calculations from one step to the next. Following that is a section titled "Interpretation," which identifies connections between the normalized mappings and idealized, integrable, fictitious Hamiltonian models. A final section contains closing comments, some of which may - but probably will not - preview work to be done later.

My reasons for writing this memo and its predecessor have already been expressed. [1] To them can be added this: "black box code" encourages users to proceed with little or no understanding of what it does or how it operates. So far, CHEF has avoided this trap admirably by failing to attract potential users. However, we reached a watershed last year: even I now have difficulty following the software through its maze of operations. Extensions to CHEF's *physics* functionalities, software upgrades, and even simple maintenance are becoming more difficult than they should. I hope these memos will mark parts of the maze for easier navigation in the future.

Despite appearances to the contrary, I tried to include no (or very little) more than the minimum needed to understand what CHEF's nonlinear analysis modules do.[1] As with the first memo, material has been lifted - and modified - from *Intermediate Classical Dynamics* (ICD) [2], old technical memos, seminar viewgraphs, and lecture notes. Finally, for a reason I do not know but am willing to indulge, equation and comment labels start from where they left off in Part 1.

2 Prerequisites: notation, definitions, lemmas, and assertions

This section need not be read but should be referred to as needed. In it are telegraphically compiled prerequisites for the ones that follow. More complete introductions, explanations, and examples can be found in various sections of ICD [2]. I do not intend to reproduce them in their entirety but will extract and modify pieces indiscriminately.[2]

---

[1] To save reading time, permit me to identify what is important: Equations (39), (41), (42), (50), (52), and the pseudocode that follows Eq.(41).

[2] Excerpts from ICD and other copyrighted material are reproduced with permission of the author.

To begin, an "observable" is a function defined over phase space, and a coordinate is a special case of an observable. However, we need to expand this concept from real valued functions to complex valued ones as well; complex observables and complex coordinates are not only permitted but not quite essential but exceedingly useful for this formalism. As usual, R represents the field of real numbers and C, of complex. The symbol $\mathcal{F}$ will stand in for either R or C, when it is convenient to leave the field ambiguous.

On those few occasions when it is convenient to do so, I'll symbolize phase space by "$\mathcal{P}$." Calculations take place within localized, open subsets with coordinates written on local charts: open subsets of either R^{2n} or C^{2n}, for some n, which contain the origin.

The dimensions of C^{2n} and R^{2n} are different, but as in the first memo, we are restricted to the subset

$$\{\, (a_1, a_2, \ldots, a_{2n}) \in C^{2n} \mid \forall k : a_{k+n} = a_k^* \,\} \;\subset\; C^{2n} \;, \tag{29}$$

where index arithmetic will always be understood to proceed modulo $2n$. I shall abuse the notation, C^{2n}, by henceforth assuming this restriction everywhere without further comment. There are other subtleties, such as angle coordinates not being defined at zero action, so that $\mathcal{P}$ technically must be "punctured" by removing the origin and adding it back as a "special point." I shall ignore that, and all similar embarassments, because they make it awkward to write intuitively simple sentences like, "The origin must be a fixed point of the mapping."

2.1 Polynomial mappings

The subjects of our attention are smooth invertible mappings,

$$\mathcal{M} : \mathcal{F}^{2n} \to \mathcal{F}^{2n} \;,$$

where notational abuses have already been noted (and will never again be mentioned). More specifically, we are primarily interested in the jet[3] representatives of such mappings: polynomial maps, i.e. mappings whose components are polynomials.[4] Even then, we shall restrict ourselves only to those that map the origin into itself.

No good, universally accepted notation has been invented for writing about all this. For openers, I'll switch freely between several ways of representing coordinates and between using N or $2n$ to represent the dimension of phase space,

either as an ordered tuple:	e.g.	$(u_1, u_2, \ldots, u_N)$	$\in \mathcal{F}^N$
or as an underlined symbol:	e.g.	$\underline{u}$	$\in \mathcal{F}^N$
or as an indexed set of variables:	e.g.	$\{\, u_k \mid k = 1..2n \,\}$, or simply $\{u_k\}$	$\in \mathcal{F}^N$

depending on convenience within a current context. We give a special designation, "$\Phi_{\underline{\sigma}}$," indexed by exponents, to a monomial with coefficient one.

$$\Phi_{\underline{\sigma}} : \mathcal{F}^N \to \mathcal{F}, \quad \{u_k\} \mapsto \prod_{k=1}^{N} u_k{}^{\sigma_k}, \quad \text{with all } \sigma_k \in \{0, 1, 2, 3, \ldots\} \;.$$

The degree of this term is the sum of the exponents; the degree of a general polynomial is the maximum degree of any of its terms.[5] If polynomials are viewed as forming a vector space over $\mathcal{F}$, then the subset, $\cup_{\underline{\sigma}} \{\Phi_{\underline{\sigma}}\}$, span the subspace in

---

[3] I cannot remember where I first read this, but in the theory of Lie-Backlund transformations a "jet" is an equivalence class of mappings all of whose derivatives are identical, up to a specified order, at one given point. It is, if you will, a generalization of the concept "tangent space." More on this can be found in the 1995 revision of the MXYZPTLK User's Guide. [3]

[4] I had thought that "polynomials" were functions taking one argument, $P : F \to F$, and used "multinomials" for those taking several arguments, $P : F^N \to F$. Evidently, correct mathematical terminology employs the word "polynomial" in both cases.

[5] I shall use the words "degree" and "order" interchangeably, as if they had the same meaning. This is not so in other variants of perturbation theory, but for us the order parameter is always linked to some power of the coordinates' scale.

which we shall work.

$$P : \mathcal{F}^N \to \mathcal{F}, \quad P = \sum_{\{\underline{\sigma} \,|\, 0 < \Sigma_k \sigma_k \leq M\}} c_{\underline{\sigma}} \Phi_{\underline{\sigma}}, \quad \{c_{\underline{\sigma}}\} \subset \mathcal{F} \ .$$

The positive integer M is the degree of the polynomial. **NOTE:** We disallow the constant term, $\underline{\sigma} = \underline{0}$, to enforce the constraint $P(\underline{0}) = 0$.

Next, we need a notation for projecting $c_{\underline{\sigma}}$ out of P. Again thinking of P as belonging to a vector space, we can employ an inner product notation, such as

$$c_{\underline{\sigma}} = (\Phi_{\underline{\sigma}}, P)$$

Many accelerator physicists have adopted an abuse of Dirac notation,

$$c_{\underline{\sigma}} = \langle \underline{\sigma} | P \rangle$$

that is frequently convenient though sometimes ambiguous. Should we need to filter out more than the coefficient, we can define an indexed set of projection operators.

$$\Pi_{\underline{\sigma}} : \ \mathcal{F}^{\mathcal{F}^N} \to \mathcal{F}^{\mathcal{F}^N}, \quad P \mapsto c_{\underline{\sigma}} \Phi_{\underline{\sigma}}$$

To project *components* of a mapping we shall use a notation that anticipates its computer implmentation. If $f : \ \mathcal{F}^N \to \mathcal{F}^N$ is a mapping, then $f[i]$ will represent its ith component. That is,

if $f : \ \mathcal{F}^N \to \mathcal{F}^N, \ (u_1, u_2, \ldots, u_N) \mapsto (u'_1, u'_2, \ldots, u'_N)$ **then** $f[i] : \ \mathcal{F}^N \to \mathcal{F}, \ (u_1, u_2, \ldots, u_N) \mapsto u'_i$.

Finally, we tacitly acknowledge that all functions inherit the algebraic properties of their image spaces: for example, for $f, g : \ \mathcal{F}^N \to \mathcal{F}$ and $c \in \mathcal{F}$,

$$h = f + g \quad \Leftrightarrow \quad h(\underline{u}) = f(\underline{u}) + g(\underline{u})$$
$$h = f \cdot g \quad \Leftrightarrow \quad h(\underline{u}) = f(\underline{u}) \cdot g(\underline{u})$$
$$h = c \cdot f = cf \quad \Leftrightarrow \quad h(\underline{u}) = c f(\underline{u})$$

and so forth. To this we must include the composition algebra, *when the range of one function matches the domain of another*: e.g. for $g : \ \mathcal{F}^N \to \mathcal{F}$ and $f : \ \mathcal{F} \to Z$,

$$f \circ g : \ \mathcal{F}^N \to Z, \quad \underline{u} \mapsto f(g(\underline{u})) \ .$$

Notice that polynomial maps are closed under composition: i.e. if f and g are polynomial maps, then so is $f \circ g$.

In everything that follows, the symbol $\mathcal{I}$ will represent the identity mapping in whatever space happens to be the current context.

for all $u \in R : \ \mathcal{I}(u) = u;$ **for all** $u \in \mathcal{F} : \ \mathcal{I}(u) = u;$ **for all** $\underline{u} \in \mathcal{F}^N : \ \mathcal{I}(\underline{u}) = \underline{u};$ and so on ...

COMMENT:

23. All this notational fussbudgetry is not meant to be either pedantic or didactic.[6] Considering what lies ahead, this seems a little like straining out gnats and (later) swallowing camels [4]. Nonetheless, when the time comes to translate equations into lines of code – or, even worse, to design C++ classes with useful attributes and behaviors – it helps to have identified the categories of objects that operators act upon and are expected to return.

---

[6]To paraphrase from Woody Allen's *Broadway Danny Rose.*

2.2 Function algebras

Collected here are a few lemmas that will be needed on the algebras of mappings. All but one are too trivial to bother writing their proofs.

LEMMA: Homogeneous functions, $\{ f : \mathcal{F}^N \mapsto \mathcal{F} \mid f(\lambda \underline{u}) = \lambda^p f(\underline{u}), \text{ for all } \lambda, u \in \mathcal{F}, \text{ for some } p \in Z^+ \}$, form a graded algebra under multiplication, with the multiplier's exponent as the grade.

$$\deg(f \cdot g) = \deg(f) + \deg(g)$$

LEMMA: Polynomial mappings form a graded algebra under addition and composition. A mapping's grade is its degree, which is propagated through algebraic operations as follows:

$$\deg(f + g) = \max(\deg(f), \deg(g)) \quad \textbf{and} \quad \deg(f \circ g) = \deg(f) \cdot \deg(g) \ .$$

In particular, if $\deg(g) = 1$, then $\deg(f \circ g) = \deg(f)$.

That last, innocuous sentence makes possible the procedure that is the subject of this memo. These first two lemmas will have increased importance for the material in the next one.

LEMMA: Composition (for both functions and mappings) is right distributive over addition. It is also left distributive for linear maps. That is,

$$\textbf{for all } f, g, h: \quad (g + h) \circ f = (g \circ f) + (h \circ f)$$
$$f \circ (g + h) = (f \circ g) + (g \circ h) \ \textbf{if and only if} \ \ f \text{ is linear.}$$

The next lemma is non-intuitive and important enough to merit writing its proof, albeit as simple as the others.

LEMMA: Let $Q^{(p)}$ be a (specific) homogenous polynomial map of degree $p > 1$, and let $O^{[p+1]}$ represent a generic (possibly infinite) polynomial map whose *lowest* order term is of degree $p + 1$. Then, the composite inverse of $\mathcal{I} + Q^{(p)} + O^{[p+1]}$ is,

$$(\mathcal{I} + Q^{(p)} + O^{[p+1]})^{-1} = \mathcal{I} - Q^{(p)} + O^{[p+1]} \ .$$

PROOF: This proof is nothing more than a simple calculation that uses the previous lemma. To improve its readability, I shall replace "$+ O^{[p+1]}$" with "$+ \cdots$."

$$\textbf{for all } \underline{u}: \ [[\mathcal{I} - Q^{(p)} + \cdots] \circ [\mathcal{I} + Q^{(p)} + \cdots]](\underline{u}) \ = \ [\mathcal{I} + Q^{(p)} + \cdots - Q^{(p)} \circ [\mathcal{I} + Q^{(p)} + \cdots] + \cdots](\underline{u})$$
$$= \ \underline{u} + Q^{(p)}(\underline{u}) - Q^{(p)}(\underline{u} + Q^{(p)}(\underline{u}) + \cdots) + \cdots$$
$$= \ \underline{u} + Q^{(p)}(\underline{u}) - Q^{(p)}(\underline{u}) + \cdots$$
$$= \ \underline{u} + \cdots$$

Written functionally,

$$[\mathcal{I} - Q^{(p)} + \cdots] \circ [\mathcal{I} + Q^{(p)} + \cdots] \ = \ \mathcal{I} + \cdots \ .$$

$\square$

<u>COMMENTS</u>:

24.   That $\mathcal{I} + Q^{(p)} + \cdots$ is a finite polynomial map does not mean that its inverse is one. For example, on the real axis, the inverse of $f : x \mapsto x + x^2$ is $f^{-1} : x \mapsto (\sqrt{4x+1} - 1)/2 = x - x^2 + \cdots$. More could be written about this – e.g. branch cuts, domains – but why bother? One point to be noted *en passant* is that CHEF's mxyzptlk library contains a method to compute the exact composite inverses of jets automatically.

25.   Nothing within this memo has anything to do with convergence. Estimating domains of convergence of the (possibly asymptotic) series soon to be written is a research area linked with the existence, strengths, and locations of resonances. I have no intention of stepping into it here; just establishing a sensible topology would by itself take us too far astray. Treat all operations as formal, as though our purpose does not go beyond the minimum needed to understand CHEF's source code. (Oh, wait: it doesn't!)

2.3   Action-angle and Weyl coordinates

We shall need the differential forms connecting complex Weyl (normal form) coordinates and action-angle coordinates from Eq.(13) of the previous memo or from ICD [2, page 24, Eq.(2.9)].

$$\textbf{for } j = 1..n: \quad a_j = i\sqrt{I_j}\, e^{-i\varphi_j} \implies da_j/a_j = dI_j/2I_j - i\, d\varphi_j \ .$$

Since we are not being rigorous I won't bother pointing out subtleties, such as the fact that $d\varphi_j$ is globally well defined but φ_j is not. The corresponding dual forms will be even more useful:[7]

$$
\begin{aligned}
\frac{\partial}{\partial I_j} &= \frac{1}{2}\left(\frac{1}{a_j^*}\frac{\partial}{\partial a_j} + \frac{1}{a_j}\frac{\partial}{\partial a_j^*} \right) \\[2em]
\frac{\partial}{\partial \varphi_j} &= i\left(a_j^* \frac{\partial}{\partial a_j^*} - a_j \frac{\partial}{\partial a_j} \right) \ ,
\end{aligned}
\tag{30}
$$

the inverses of which are,

$$\frac{\partial}{\partial a_j} = a_j^* \frac{\partial}{\partial I_j} + \frac{i}{2a_j}\frac{\partial}{\partial \varphi_j} \quad \text{or, if you prefer,} \quad a_j\frac{\partial}{\partial a_j} = I_j\frac{\partial}{\partial I_j} + \frac{i}{2}\frac{\partial}{\partial \varphi_j} \ .$$

2.4   Lie (derivative) operators

Much of this section will be lifted directly from the pages of ICD [2]. We begin with the notion of a "vector field" – also known as "system," "dynamical system," "dynamic," or "ordinary differential equation." For our purposes, this is an array of observables, usually interpreted as the (local) components of a differential equation.

$$\underline{\dot{z}} = d\underline{z}/d\theta \equiv \underline{v}(\underline{z}\,[,\theta])$$

The symbol θ is used as the orbit parameter in anticipation of defining the one-turn map,

$$\mathcal{M} : \underline{z}(\theta) \mapsto \underline{z}(\theta + 2\pi) \ .$$

---

[7]These appeared as Problem 2.4 (page 97) of ICD [2].

If $\underline{v}$ has no explicit dependence on θ, then the system is "autonomous."

Lie Derivative.   Given any observable, g, and local representation, $\underline{v}$, of the vector field, we define the linear operator, $L_{\underline{v}}$, as[8]

$$L_{\underline{v}}g = h, \quad \text{where} \quad h(\underline{z}) = \underline{v}(\underline{z}) \cdot \frac{\partial g(\underline{z})}{\partial \underline{z}} \ .$$

Depending on author and context, $L_{\underline{v}}g$ is called the "Lie derivative," "Lie operator," or simply a "[tangent] vector [field]." We note in passing that (a) $h = L_{\underline{v}}g$ is again an observable, at least locally, so that (a.1) $L_{\underline{v}}$ indeed acts like a linear operator on the space of observables (treated as a vector space) and (a.2) it can be applied repeatedly (powers of $L_{\underline{v}}$ are well defined) and (b) $L_{\underline{v}}$ is a geometric object[9], i.e. if a coordinate transformation is effected,

$$\mathcal{T}: \ \underline{z} \mapsto \underline{z}' \ \textbf{and} \ \underline{v}(\underline{z}) \mapsto \underline{v}'(\underline{z}') \ ,$$

then

$$\underline{v}(\underline{z}) \cdot \frac{\partial}{\partial \underline{z}} = \underline{v}'(\underline{z}') \cdot \frac{\partial}{\partial \underline{z}'} \ .$$

The Lie derivative has a natural interpretation as either a directional derivative or a "time" derivative, through the application of the chain rule.[10]

$$\frac{dg}{d\theta} = \frac{d\underline{z}}{d\theta} \cdot \frac{\partial g}{\partial \underline{z}} = \underline{v} \cdot \underline{\partial} g = L_{\underline{v}}g \ . \tag{31}$$

Exponential maps.   If $L_{\underline{v}}$ is autonomous – or if autonomy is forced by extending phase space – and g itself is independent of the orbit parameter, then as an operator equation on the linear vector space of observables, Eq.(31) has the immediate (and more than formal) solution.

$$g_\theta = e^{\theta L_{\underline{v}}} g_0$$

We write g_θ with a subscript to emphasize that this represents a one-parameter family of observables. Depending on context this is called either an exponential map, the (classical) time evolution operator, the Heisenberg picture, or simply Taylor's theorem. In particular, replacing g_0 with the identity provides an expression for what we would call the one-turn mapping.

$$\mathcal{M} = e^{2\pi L_{\underline{v}}} \mathcal{I}$$

Earlier chapters of ICD [2, pages 63ff] contain examples of how this generates the orbits of simple, exactly solvable dynamical systems. In the present context, it will be used in Section 4 to connect underlying Hamiltonian models with mappings.

Hamiltonian dynamics.   A Hamiltonian dynamic is obtained by rotating a gradient vector field by $\pi/2$ along a specific direction in phase space.[11] Let $H : \mathcal{P} \to R$ be any observable. An associated Hamiltonian dynamic[12] has *the components*,

$$\underline{v} \ = \ d\underline{z}/d\theta \ = \ \underline{\underline{J}} \cdot \underline{\partial} H \ , \quad \text{that is, for } i = 1..2n, \ v_i = \sum_{k=1}^{2n} J_{ik} \frac{\partial H}{\partial z_k} \ ,$$

---

[8]This is standard notation among mathematicians and most mathematical physicists.

[9]This is like writing that $\vec{v}$ is a vector whose components are v_i. The operators $\partial/\partial z_i$ comprise the "natural" basis for the tangent spaces attached to points of $\mathcal{P}$. For an extended discussion of this, see ICD [2] or, better, any good book on modern differential geometry.

[10]This equation provides the simplest argument that $L_{\underline{v}}$ *must* be a geometric, coordinate-independent entity, as $dg/d\theta$ cannot depend on the coordinates used to evaluate it. In fact, Eq.(31) is the starting point for more exact, formal treatments of the theory of tangent spaces and tangent bundles. From it is *defined* what one literally means by the tangent space attached to a manifold at a point.

[11]This is stated loosely. The concept of "angle" has no natural definition in a space that does not possess a metric.

[12]What I've written here is a simplified treatment, specialized to global canonical coordinates.

where $\underline{z}$ are canonical coordinates over $\mathcal{P}$ ordered as in the first memo, so that

$$\underline{\underline{J}} \equiv \begin{pmatrix} \underline{\underline{0}} & \underline{\underline{1}} \\ -\underline{\underline{1}} & \underline{\underline{0}} \end{pmatrix} \ ,$$

Mathematicians symbolize the vector field with these components as X_H; in his seminal 1982 lectures at the first USPAS session [5], Alex Dragt introduced a colon notation into accelerator physics to represent the same thing (apart from a sign).

$$X_H \ = \ L_{\underline{v}} \ = \ -:H: \ = \ \underline{v} \cdot \underline{\partial} \ = \ -\underline{\partial} H \cdot \underline{\underline{J}} \cdot \underline{\partial}$$

If we separate $\underline{z}$ into "position" and "momentum" sectors in coordinate systems of interest, the components take on readily recognizable forms.[13]

"Cartesian" coordinates: $\quad \underline{z} = \begin{pmatrix} \underline{x} \\ \underline{p} \end{pmatrix}, \quad \underline{\dot{x}} = \partial H/\partial \underline{p}, \quad \underline{\dot{p}} = -\partial H/\partial \underline{x},$

action-angle coordinates: $\quad \underline{z} = \begin{pmatrix} \underline{\varphi} \\ \underline{I} \end{pmatrix}, \quad \underline{\dot{\varphi}} = \partial H/\partial \underline{I}, \quad \underline{\dot{I}} = -\partial H/\partial \underline{\varphi},$

normal (Weyl) coordinates: $\quad \underline{z} = \begin{pmatrix} \underline{a} \\ \underline{a}^* \end{pmatrix}, \quad \underline{\dot{a}} = -i\partial H/\partial \underline{a}^*, \quad \underline{\dot{a}}^* = i\partial H/\partial \underline{a},$

Because X_H is a geometric object[14] - a vector, independent of the coordinate system - we can equate,

$$\begin{aligned} X_H \ &= \ \partial H/\partial \underline{p} \cdot \partial/\partial \underline{x} - \partial H/\partial \underline{x} \cdot \partial/\partial \underline{p} \\ &= \ \partial H/\partial \underline{I} \cdot \partial/\partial \underline{\varphi} - \partial H/\partial \underline{\varphi} \cdot \partial/\partial \underline{I} \\ &= \ 2\Im(\partial H/\partial \underline{a}^* \cdot \partial/\partial \underline{a}) \ . \end{aligned} \quad (32)$$

A fundamental property is that *Hamiltonian fields are closed under commutation*, with

$$\textbf{for all } f,g : \ [X_f, X_g] = X_{\{g,f\}} \ ,$$

where $\{\cdot,\cdot\}$ indicates a Poisson bracket.[15]

2.5 Integrable Hamiltonian models

A not totally necessary step in map analysis is classifying the one-turn mapping as being "equivalent to" one generated by a member of a collection of (preferably) integrable Hamiltonians. The first of these is the linear model,

$$H \ = \ \underline{v} \cdot \underline{I} \ = \ \sum_{j=1}^{n} v_j I_j, \quad \text{and} \quad X_H \ = \ \underline{v} \cdot \frac{\partial}{\partial \underline{\varphi}} \ = \ \sum_{j=1}^{n} v_j \frac{\partial}{\partial \varphi_j} \ .$$

At this level, mappings are (linearly) equivalent if they have the same tunes. One level up are the shearing Hamiltonians, which contain nonlinear terms in the action coordinate.

$$H \ = \ \underline{v} \cdot \underline{I} + H_s(\underline{I}), \quad \text{and} \quad X_H \ = \ \left(\underline{v} + \frac{\partial H_s}{\partial \underline{I}} \right) \cdot \frac{\partial}{\partial \underline{\varphi}} \ \equiv \ \underline{v}(\underline{I}) \cdot \frac{\partial}{\partial \underline{\varphi}} \quad (33)$$

---

[13]For an explanation of the extra factor of i appearing in the third equation, see ICD [2, page 133].

[14]Again, see ICD [2, page 59, Eq.(2.54); page 130, Eq.(3.46)].

[15]This defines a morphism between the commutator algebra of Lie operators of Hamiltonian dynamics and the Poisson bracket algebra of observables. For a quick derivation and discussion, see ICD [2, page 108]; for further discussion, read Abraham and Marsden [6].

These provide representative models possessing amplitude-dependent tunes. The final category we shall consider contains integrable resonance models. First the autonomous,

$$
\begin{aligned}
H & = \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) + H_{r,s}(\underline{I}) \sin(\underline{m}^T \cdot \underline{\varphi}) + H_{r,c}(\underline{I}) \cos(\underline{m}^T \cdot \underline{\varphi}) \\
& = \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) + H_r(\underline{I}) \cos(\underline{m}^T \cdot \underline{\varphi} + \eta(\underline{I})) \quad,
\end{aligned}
\tag{34}
$$

where $\underline{m}$ is an array of integers. To this is added the periodic version,

$$
H = \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) + H_{r,s}(\underline{I}) \sin(\underline{m}^T \cdot \underline{\varphi} + n\theta) + H_{r,c}(\underline{I}) \cos(\underline{m}^T \cdot \underline{\varphi} + n\theta) \quad,
\tag{35}
$$

were n is an arbitrary integer, not half the dimension of $\mathcal{P}$. By a well known sequence of manipulations, Eq.(35) can be transformed into Eq.(34).[16] That this is so effectively proves its integrability.

The resonance models written above are specializations of the more generic,

$$
H = \underline{\nu} \cdot \underline{I} + H_s(\underline{I}) + H_r(\underline{I}) F(\underline{m}^T \cdot \underline{\varphi}) \quad,
$$

where F is a (not arbitrary) periodic function. We shall shun the more general possibilities and be content with a single trigonometric function.

COMMENTS:

26. The attraction of integrable models is that the asymptotic global behavior of their orbits is (semi-)analytically calculable. Thus, their stable regions of phase space can be characterized unambiguously without resorting to tracking. Further, they are characterized by a small number of control parameters, such as tunes and resonance strengths.

27. One way of classifying or "interpreting" a symplectic mapping is arguing that it is equivalent to the exponential map generated by one of these models by constructing a coordinate transformation that converts one into the other. That is,

$$
\mathcal{M} \simeq e^{2\pi X_H} \, \mathcal{I}, \ \text{ or perhaps } \ \mathcal{M} \simeq \left(\prod_i e^{2\pi X_{H_i}} \right) \mathcal{I} \quad,
\tag{36}
$$

where the latter product is ordered and the X_{H_i} do not necessarily commute.[17] Such transformations almost never exist globally, so the argument proceeds to its conclusion by convincing the reader (or listener) - by whatever means are at one's disposal - that a usually infinite number of (higher order) terms can be ignored. Failing that, one appends the phrase "at this order" to the end of the argument and hopes for the best. The preceding sentences could initiate a long discussion that I do not intend to enter. Instead, we shall turn to constructing the required transformation(s).

3 Iterative sequence of transformations

With these preliminaries out of the way, we are ready to describe the perturbative procedure called "map analysis." Its first, linear step was already explained in the previous memo, with all the details needed to assure that normalization preserves the Poincaré invariants, including the (polarized) phase space volume. Subsequent, nonlinear steps should maintain that objective while iteratively bringing the original symplectic mapping into normal form.

---

[16]This procedure is described in gory detail on pages 205-211 of ICD [2].

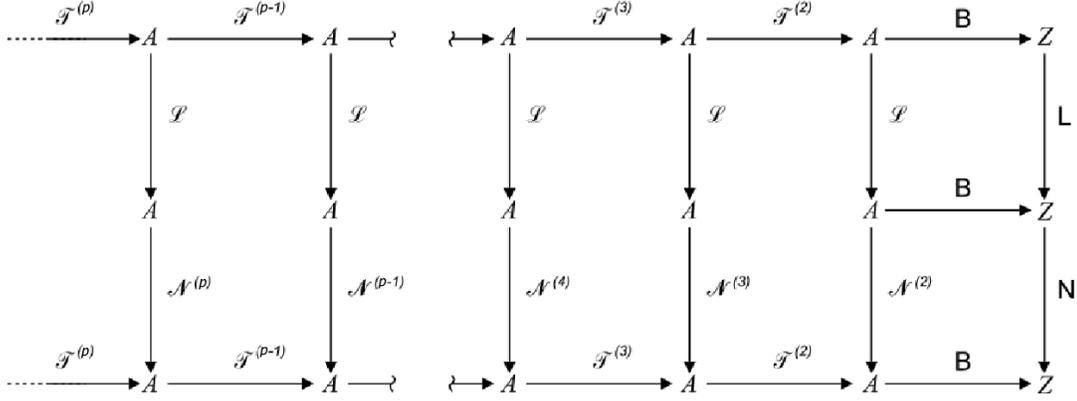[17]This is, of course, the goal of MARYLIE.

Figure 1: Ladder of normal form transformations. Each step's "order," or "degree," is annotated as a parenthetical super-script. The rightmost composition, $N \circ L$, is the (one-turn) nonlinear mapping expressed in the original phase space. The square in the upper right hand corner represents reduction to linear normal form.

3.1 Transformative tandem

Shown as a commutative diagram in Figure 1 is the iterative ladder of transformations that reduce a nonlinear map to its normal form. Each step's "order," or "degree," is annotated as a parenthetical superscript. The square in the upper right hand corner represents the linear piece described in the previous memo. The rightmost composition, $N \circ L = M$, is the (one-turn) nonlinear mapping expressed in the original phase space coordinates. Such a factorization is "always" possible: simply set $N \equiv M \circ L^{-1}$. As a result, the linear part of N is the identity. This simple and exceedingly useful trick is carried over throughout the procedure: the linear part of all $\mathcal{N}^{(p)}$, as well as all $\mathcal{T}^{(p)}$, will be the identity. In a slight abuse of the word "near," such mappings are frequently called "near identity transformations," but again I emphasize that nothing in this memo has anything to do with convergence.

> **The objective** on the $(p-1)^{\text{th}}$ step of this ladder is to absorb as many terms as possible of degree $p-1$ from the mapping $\mathcal{N}^{(p-1)}$ into the transformation $\mathcal{T}^{(p-1)}$ so that they will not appear in $\mathcal{N}^{(p)}$.

That is the most important sentence in the entire memo; everything else just describes manipulations for carrying it out. Meditate on the words while staring at Figure 1 until both become irreversibly imprinted. After that happens, proceed to read the definition of $\mathcal{N}^{(p)}$ from the diagram by traversing the $(p-1)^{\text{th}}$ step clockwise.

$$\mathcal{N}^{(p)} = (\mathcal{T}^{(p-1)})^{-1} \circ \mathcal{N}^{(p-1)} \circ L \circ \mathcal{T}^{(p-1)} \circ L^{-1} \tag{37}$$

$\mathcal{N}^{(p-1)}$ and, of course, L are already known; $\mathcal{T}^{(p-1)}$ must now be constructed so as to achieve **the objective** for $\mathcal{N}^{(p)}$. To

that end, we choose $\mathcal{T}^{(p-1)}$ to be,

$$\mathcal{T}^{(p-1)} = \mathcal{I} + \mathcal{T}^{(p-1)}_{p-1} + \cdots \ , \tag{38}$$

where the lowest nonlinear terms form a homogeneous polynomial map, $\mathcal{T}^{(p-1)}_{p-1}$, of degree $p-1$. In principle, at this point we could take those as the only terms, but we shall leave open the possibility of later adding more, of degree $\geq p$, which we are temporarily buried in the dots.

Using an lemma from the previous section,

$$(\mathcal{T}^{(p-1)})^{-1} = \mathcal{I} - \mathcal{T}^{(p-1)}_{p-1} + \cdots \ .$$

Further, we expose $\mathcal{N}^{(p-1)}$'s series of homogeneous polynomial maps,

$$\mathcal{N}^{(p-1)} = \mathcal{I} + \mathcal{N}^{(p-1)}_2 + \mathcal{N}^{(p-1)}_3 + \cdots + \mathcal{N}^{(p-1)}_{p-1} + \mathcal{N}^{(p-1)}_p + \cdots \ ,$$

of which we are primarily interested, at this step, in $\mathcal{N}^{(p-1)}_{p-1}$. Inserting these expressions into Eq.(37) isolates the terms of (current) interest.

$$\mathcal{N}^{(p)} = (\mathcal{I} - \mathcal{T}^{(p-1)}_{p-1} + \cdots) \circ (\mathcal{I} + \cdots + \mathcal{N}^{(p-1)}_{p-1} + \cdots) \circ \mathcal{L} \circ (\mathcal{I} + \mathcal{T}^{(p-1)}_{p-1} + \cdots) \circ \mathcal{L}^{-1}$$

Now using the distributive lemma, and the (crucial) fact that $\mathcal{L}$ is linear, we can rework the three rightmost factors,

$$\mathcal{L} \circ (\mathcal{I} + \mathcal{T}^{(p-1)}_{p-1} + \cdots) \circ \mathcal{L}^{-1} = \mathcal{I} + \mathcal{L} \circ \mathcal{T}^{(p-1)}_{p-1} \circ \mathcal{L}^{-1} + \cdots \ ,$$

to produce

$$\begin{aligned}
\mathcal{N}^{(p)} &= (\mathcal{I} - \mathcal{T}^{(p-1)}_{p-1} + \cdots) \circ (\mathcal{I} + \cdots + \mathcal{N}^{(p-1)}_{p-1} + \cdots) \circ (\mathcal{I} + \mathcal{L} \circ \mathcal{T}^{(p-1)}_{p-1} \circ \mathcal{L}^{-1} + \cdots) \\
&= \mathcal{I} + \cdots + \mathcal{N}^{(p-1)}_{p-1} - (\mathcal{T}^{(p-1)}_{p-1} - \mathcal{L} \circ \mathcal{T}^{(p-1)}_{p-1} \circ \mathcal{L}^{-1}) + \cdots \ .
\end{aligned}$$

The final expression is obtained by continuing to apply the right distributive property and ruthlessly shoving higher order terms into the dots as they appear. Projecting out from $\mathcal{N}^{(p)}$ the homogenous map of degree $p-1$ provides the following evocative result.

$$\mathcal{N}^{(p)}_{p-1} = \mathcal{N}^{(p-1)}_{p-1} - (\mathcal{T}^{(p-1)}_{p-1} - \mathcal{L} \circ \mathcal{T}^{(p-1)}_{p-1} \circ \mathcal{L}^{-1}) \tag{39}$$

For reasons definitely not worth going into, this is called the "homological equation."

<u>COMMENTS</u>:

28.   While not yet completely obvious, the nature of this tandem will mean that, once it is converted into an algorithm, the algorithm is the same at all orders. Thus, written once, it can be applied iteratively. More than anything else, that is what originally made this entire procedure so attractive.

29.   However, it was not the first such breakthrough. Poincaré anticipated this technique but, not possessing computers, did not push it beyond the first two steps.[18] In the 1960's, André Deprit [7] developed a corresponding iterative procedure for celestial mechanics that operated in the realm of autonomous Hamiltonian dynamics not necessarily expressed as polynomials.[19] Deprit's algorithm was my first exposure to nonlinear perturbative techniques that could be automated to

---

[18]I once had the reference, have lost it, but shall search for it.

[19]An extension that handles non-Hamiltonians as well is fully derived in Section 5.2 of ICD [2, pages 253ff].

any order. It was the first significant advance in this area since Poincaré, and its elegance contrasted sharply with the manipulations necessary to get results even to second order[20] using the generating function techniques emphasized in the overly venerated text of Goldstein [12].

30. I once had the privilege, in the late 1980's, of spending a day with Deprit at the Naval Research Laboratory. He informed me that even within his own field of celestial mechanics his revolutionary accomplishment was not trusted. He had been applying it to establish the stability of satellite orbits, yet those higher in his administrative food chain insisted on continually "benchmarking" his answers against numerical integration. André Deprit died in November, 2006.

3.2 Solving the homological equation

We have now reached the heart of the matter: solving Eq.(39). Because $\mathcal{L}$ is linear this can be handled one monomial term at a time. To ease the notation, I am temporarily dropping p and $(p-1)$ subscripts and superscripts, with the understanding that we are positioned on step $p-1$ of the ladder. Expand,

$$\mathcal{T} = \sum_{\underline{\sigma}} t_{\underline{\sigma}} \Phi_{\underline{\sigma}}$$

and consider that

$$\Phi_{\underline{\sigma}} \circ \mathcal{L}^{-1} : \quad \underline{a} \mapsto \Phi_{\underline{\sigma}}(\underline{\underline{\Lambda}}^{-1}\underline{a}) = \prod_{k=1}^{2n}((e^{i\mu_k}a_k)^{\sigma_k} = \exp\left(i\sum_{k=1}^{2n}\sigma_k\mu_k\right)\Phi_{\underline{\sigma}}(\underline{a}) \ ,$$

where it is understood that $\mu_{k+n} = -\mu_k$, so that

$$\exp\left(i\sum_{k=1}^{2n}\sigma_k\mu_k\right) = \exp\left(i\sum_{k=1}^{n}(\sigma_k - \sigma_{k+n})\mu_k\right) \ .$$

Mathematical notation becomes a little awkward at this point. The remaining left operation by $\mathcal{L}$ is easier to explain in words: "the jth component of $\mathcal{T}$ is multiplied by $\exp(-i\mu_j)$." Symbolically, I want to write something simple, like,

$$\textbf{for all } j : \quad \textbf{for all } \underline{\sigma} : \quad \mathcal{L} \circ \Phi_{\underline{\sigma}} \circ \mathcal{L}^{-1} = \exp\left(i\sum_{k=1}^{2n}(\sigma_k - \delta_{j,k})\mu_k\right)\Phi_{\underline{\sigma}} \ ,$$

with $\delta_{j,k}$ being the usual Kronecker delta. While this captures the essence of what must happen, it is wrong on two counts: (1) j does not appear on the lefthand side of the equality, and (2) the domain of $\mathcal{L}$ is the space of observables on C^{2n} – i.e. functions mapping $C^{2n} \to C^{2n}$ – while the argument presented to it is a function mapping $C^{2n} \to C$. It is more correct to use an inner product to write something like,

$$\textbf{for all } j : \quad \textbf{for all } \underline{\sigma} : \quad (\Phi_{\underline{\sigma}}, (\mathcal{L} \circ \mathcal{T} \circ \mathcal{L}^{-1})[j]) = \exp\left(i\sum_{k=1}^{2n}(\sigma_k - \delta_{j,k})\mu_k\right)(\Phi_{\underline{\sigma}}, \mathcal{T}[j]) \ ,$$

---

[20]For example, the technical memo of Cole [8], though contemporaneous with Deprit's work, was considered "state of the art" at the time among accelerator physicists. In those days this was frequently called a "Moser transformation," at least at Fermilab, presumably because Moser had used such transformations in his proof of the KAM theorem. My own small contribution was to realize that Deprit's algorithm could be modified slightly to include periodic Hamiltonian systems [9] [10] [2]. It had all the impact of a gnat landing on a wet water buffalo and was rapidly superseded by the mapping techniques [11]. Nonetheless, case studies on simple models led to a "resonance seeding hypothesis" [2, pages 281-284] for quickly estimating dynamic apertures, an idea that had even less influence on subsequent events. Ah, the memories!

which, while less transparent, at least keeps track of the double projection taking place. Referring back to Eq.(39),

$$\textbf{for all } j: \quad \textbf{for all } \underline{\sigma}: \quad (\Phi_{\underline{\sigma}}, \ (\mathcal{T} - \mathcal{L} \circ \mathcal{T} \circ \mathcal{L}^{-1})[j] \,) \ = \ \left(1 - \exp\left(i \sum_{k=1}^{2n} (\sigma_k - \delta_{j,k}) \mu_k \right) \right) (\Phi_{\underline{\sigma}}, \ \mathcal{T}[j] \,) \ .$$

Finally, let the coefficients of $\Phi_{\underline{\sigma}}$ in the jth component of $\mathcal{N}_{p-1}^{(p-1)}$, $\mathcal{N}_{p-1}^{(p)}$ and $\mathcal{T}_{p-1}^{(p-1)}$ be $g_{j,\underline{\sigma}}^{(p-1)}$, $g_{j,\underline{\sigma}}^{(p)}$ and $t_{j,\underline{\sigma}}^{(p-1)}$ respectively. Eq.(39) then boils down to the following.

$$\textbf{for all } j: \quad \textbf{for all } \underline{\sigma}: \quad g_{j,\underline{\sigma}}^{(p)} \ = \ g_{j,\underline{\sigma}}^{(p-1)} - \left(1 - \exp\left(i \sum_{k=1}^{2n} (\sigma_k - \delta_{j,k}) \mu_k \right) \right) t_{j,\underline{\sigma}}^{(p-1)} \tag{40}$$

Our objective is to choose *as many as possible* of the $t_{j,\underline{\sigma}}^{(p-1)}$ so that $g_{j,\underline{\sigma}}^{(p)} = 0$. That is, we want to construct the transformation so as to absorb as much as possible of the nonlinear behavior, removing it from the mapping. The only terms that cannot be absorbed are those for which

$$\exp\left(i \sum_{k=1}^{2n} (\sigma_k - \delta_{j,k}) \mu_k \right) \ = 1 \ . \tag{41}$$

Peeking ahead to the third memo and indulging in a little pseudocode, this piece of the algorithm can be written:

> **for each** j:
> > **for each** $\underline{\sigma}$:
> > > **if** Eq.(41) is satisfied:
> > > > $g_{j,\underline{\sigma}}^{(p)} \ := \ g_{j,\underline{\sigma}}^{(p-1)}$
> > > > $t_{j,\underline{\sigma}}^{(p-1)} \ := 0 \quad$ *[[Could be anything, but for now make it zero.]]*
> > >
> > > **else**:
> > > > $g_{j,\underline{\sigma}}^{(p)} \ := \ 0$
> > > > $t_{j,\underline{\sigma}}^{(p-1)} \ := \ g_{j,\underline{\sigma}}^{(p-1)} / \left(1 - \exp\left(i \sum_{k=1}^{2n} (\sigma_k - \delta_{j,k}) \mu_k \right) \right)$

COMMENTS:

31. If it happens that $g_{j,\underline{\sigma}}^{(p-1)} = 0$ in Eq.(40), whether Eq.(41) is satisfied becomes a non-issue. In the language of accelerator physics, a resonance is not excited if its strength vanishes.[21]

32. In the final step of the mini-algorithm above, $t_{j,\underline{\sigma}}^{(p-1)}$ may be excessively large if the denominator is small though not exactly zero, unless the numerator is correspondingly small. Therefore, when implementing the algorithm, choices must be made as to the meaning of "small." The **if** clause must be made fuzzier, and perhaps even interactive, to accomodate this ambiguity. We shall return to this in the next memo.

33. I want to avoid reinforcing the ridiculous notion that resonances come into existence because of "small denominators." Resonances – and, to some extent, shear (see next section) – are topological barriers that prevent the construction of transformations connecting the actual system to an ideal. Systems that are not topologically equivalent cannot be mapped onto one another. This is manifested in perturbative calculations by the appearance of "small denominators," but they are *effects*, not causes.

34. The combination,

$$O: \mathcal{T} \mapsto \mathcal{T} - \mathcal{L} \circ \mathcal{T} \circ \mathcal{L}^{-1}$$

---

[21]Exploiting this one fact filled much of my second and third years at Fermilab. The activity was called "magnet shuffling."

is a linear operation on the space of mappings. Eq.(39) can be rewritten in terms of this "adjoint" operator as,

$$O\mathcal{T}_{p-1}^{(p-1)} \;=\; \mathcal{N}_{p-1}^{(p-1)} - \mathcal{N}_{p-1}^{(p)} \;.$$

The terms satisfying Eq.(41) comprise the null space of the linear, "homological operator" O. Their removal amounts to nothing more than the usual requirement for handling the null space of a linear equation[22] before solving it.

3.3   Null space: shear and resonance terms

In mathematics as in biology, what cannot be absorbed must be passed on. Two types of zeroes appear while solving the homological equation: systematic and accidental. The former lead to "shear terms" corresponding to the phenomenon of amplitude dependent tunes. The latter arise from the existence of resonances and may or may not be problematic, depending on their strengths. We shall treat them separately in the next section.

   Before doing so, it may be useful to point out that Eq.(41) does not specify a unique set of exponents. Rewrite the summation, using $\mu_{k+n} = -\mu_k$ to halve the number of (formal) terms.

$$\mu_{k+n} = -\mu_k \;\implies\; \sum_{k=1}^{2n}(\sigma_k - \delta_{j,k})\mu_k \;=\; \sum_{k=1}^{n}(\sigma_k - \delta_{j,k})\mu_k + (\sigma_{k+n} - \delta_{j,k+n})\mu_{k+n}$$

$$=\; \sum_{k=1}^{n}[(\sigma_k - \delta_{j,k}) - (\sigma_{k+n} - \delta_{j,k+n})]\mu_k$$

Thus, the accidental zeroes (resonances) satisfying Eq.(41) occur when

$$\sum_{k=1}^{n}[(\sigma_k - \delta_{j,k}) - (\sigma_{k+n} - \delta_{j,k+n})]\nu_k = 0 \qquad \text{mod } 1 \;, \tag{42}$$

where I have used $\mu_k = 2\pi\nu_k$.[23] That is, the summation on the left hand side is an integer, the normal condition for a resonance line. Obviously, there is not a one-to-one correspondence between a resonance condition and a set of exponents.

   Systematic zeroes (shear terms) correspond to those $\underline{\sigma}$ which satisfy

$$\textbf{for all } k: \;\; \sigma_k - \delta_{j,k} \;=\; \sigma_{k+n} - \delta_{j,k+n} \;, \tag{43}$$

Notice the absence of "mod 1." This is an exact equality which makes Eq.(42) valid independent of the tunes. It can be written less symmetrically, but perhaps more understandably as follows.

$$\textbf{for all } k: \;\; \sigma_k - \sigma_{k+n} \;=\; \delta_{j,k} - \delta_{j,k+n} \;=\; \begin{cases} 1, & j = k \\ -1, & j = k+n \\ 0, & \text{otherwise} \end{cases}$$

<u>COMMENTS</u>:

35.   This comment could be deferred as a programming issue, but the expression $\exp\left(i\sum_{k=1}^{2n}(\sigma_k - \delta_{j,k})\mu_k\right)$ is most easily evaluated by using the already calculated diagonal elements of the linear eigenvalue matrix, $\underline{\Lambda}$.

$$\exp\left(i\sum_{k=1}^{2n}(\sigma_k - \delta_{j,k})\mu_k\right) \;=\; \prod_{k=1}^{2n}(e^{-i\mu_k})^{\delta_{j,k}-\sigma_k} \;=\; \prod_{k=1}^{2n}\Lambda_{kk}^{\delta_{j,k}-\sigma_k} \;=\; \frac{\Lambda_{jj}}{\prod_{k=1}^{2n}\Lambda_{kk}^{\sigma_k}}$$

---

[22] Including linear differential equations, resulting in the distinction between "particular" and "general" solutions.

[23] It is regrettable that ν and v look so much alike, but we must learn to tolerate such inconveniences with patience and forbearance.

This can be simplified further. For example, because $\Lambda_{kk}\Lambda_{k+nk+n} = 1$, for all k, we can replace every σ_k with $\sigma_k - \min(\sigma_k, \sigma_{k+n})$. However, let us postpone such considerations to the indefinite future.

36. When everything is connected back to physics, the extra Kronecker deltas appearing in Eq.(43) or Eq.(42), which do not appear in a continuous Hamiltonian formalism, arise from the circumstance that nonlinearities in the Hamiltonian come from a vector potential while those in the mapping come from a magnetic field.

3.4 Symplecticity: connecting the dots

All those "$+\cdots$" terms, so far hidden from view, must now be filled in. In Eq.(38), the terms left unexpressed could, in principle, be anything; they could even be nothing. However, to preserve the Poincaré invariants – which, let us agree for the moment, is a good thing – they should be chosen so that the ladder of transformations is symplectic. One way of doing that is to generate each $\mathcal{T}^{(p-1)}$ as the exponential mapping of the Lie operator associated with a scalar observable: i.e. a "Hamiltonian." This is not the only way, but it is the way of the FIB[24] algorithm [11], which CHEF currently uses and its predecessors, mxyzptlk and AESOP, have used since 1989.

So, we begin by asserting that $\mathcal{T}^{(p-1)}$ can be written as the exponential map of a Lie operator.

$$\mathcal{T}^{(p-1)} = \exp(\mathbf{T}^{(p-1)})\,\mathcal{I}\,, \tag{44}$$
$$\text{that is,} \quad \mathcal{I} + \mathcal{T}^{(p-1)}_{p-1} + \cdots = (\mathbf{1} + \mathbf{T}^{(p-1)} + \cdots)\,\mathcal{I}\,.$$

There is no subscript on $\mathbf{T}^{(p-1)}$. The components of this operator are homogeneous polynomials of a single degree; there are no higher or lower order terms. From this we identify components:

$$\mathbf{T}^{(p-1)}[j] = \mathcal{T}^{(p-1)}_{p-1}[j]\,.$$

which is to say,

$$\mathbf{T}^{(p-1)} = \sum_{j=1}^{2n} \mathbf{T}^{(p-1)}[j]\frac{\partial}{\partial a_j} = \sum_{j=1}^{2n} \mathcal{T}^{(p-1)}_{p-1}[j]\frac{\partial}{\partial a_j}\,, \tag{45}$$

The same should be the case for $\mathcal{N}^{(p)}$, which also is expected to be symplectic.

$$\mathcal{N}^{(p)} = \exp(2\pi\mathbf{N}^{(p)})\,\mathcal{I} \tag{46}$$
$$2\pi\mathbf{N}^{(p)} = \sum_{j=1}^{2n} \mathcal{N}^{(p)}_{p-1}[j]\frac{\partial}{\partial a_j}\,, \tag{47}$$

COMMENTS:

37. These alone do not guarantee symplecticity. $\mathbf{T}^{(p-1)}$ and $\mathbf{N}^{(p)}$ must actually be the Lie operator of a single, scalar observable. Nothing written so far guarantees this. Conditions must be met, to which we shall return in the next section.

38. After constructing $\mathbf{T}^{(p-1)}$ and $\mathbf{N}^{(p)}$, Equations (44) and (46) must be used to generate the entire transformation and normal form mapping *up to the order of the final calculation* before proceeding to the next step.

39. An additional factor of 2π appears in $\mathcal{N}^{(p)}$ because the exponential is to be *interpreted as* the "time"-evolution

---

[24]Forest, Irwin, Berz. I would have preferred calling it the IF algorithm, but the seminal paper [11] has three authors, and FIB is easier to pronounce than IFB. For awhile I called it the FBI algorithm, but that seems too subservsive, and FIB fits in better with the use of Lie operators.

operator of the dynamical system, **N**, over the interval $\theta \in [0, 2\pi)$. This comment leads directly to the topic of the next section.

40. Being able to write Eq.(44) easily is one advantage of using near identity transformations.

4 Hamiltonian interpretation and/or classification

By "interpretation" I mean the attempt to associate the identified "shear terms" and "resonance terms" of the mapping – i.e. the null space of the homological operator – with shear and resonance terms that would appear in an ideal, integrable Hamiltonian model. We limit our attention to Hamiltonians that can be expanded using the set $\{\Phi_{\underline{\sigma}}\}$ as a basis: that is,

$$H = \sum_{\{\underline{\sigma}\}} h_{\underline{\sigma}} \Phi_{\underline{\sigma}} \; .$$

The basis is not complete. There are many Hamiltonians that cannot be so expressed: e.g. $H = I^2 \cos(7\varphi)$, which is not analytic at the origin. Happily, we won't need them. The associated vector field – aka Lie derivative – is expanded correspondingly.

$$X_H = \sum_{\{\underline{\sigma}\}} h_{\underline{\sigma}} X_{\Phi_{\underline{\sigma}}}$$

Our objective now is to construct the coefficients, $h_{\underline{\sigma}}$, from those appearing in the mappings, $\mathcal{N}^{(p)}$. To peek ahead to the answers, see Eq.(50) and Eq.(52).

4.1 Shear terms

Start from Equations (46) and (47). Consider now the appearance of a shear term, i.e. one satisfying Eq.(43). The components of the Lie operator at a given order

$$\mathbf{N} = \sum_{j=1}^{2n} \mathcal{N}[j] \frac{\partial}{\partial a_j} = \sum_{j=1}^{n} \mathcal{N}[j] \frac{\partial}{\partial a_j} + \mathcal{N}[j+n] \frac{\partial}{\partial a_j^*} \tag{48}$$

will contain shear terms like the following.

$$\mathcal{N}[j] \quad \textbf{contains} \quad g_j \prod_{k=1}^{2n} a_k^{\sigma_k} = g_j \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} (a_k a_k^*)^{\sigma_k} \right) a_j^{\sigma_j} a_j^{*\,\sigma_j - 1}$$

$$= g_j \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) a_j$$

I have temporarily dropped the explicit "p" and "$p-1$" subscripts and superscripts which indicate our location on the transformative ladder. It is understood that all operations in this section occur within a single step. A similar result holds for $\mathcal{N}[j+n]$.

$$\mathcal{N}[j+n] \quad \textbf{contains} \quad g_j^* \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) a_j^*$$

Putting these pieces into Eq.(48) and using Eq.(30) gives us the final result.

$$\mathbf{N} \ \text{contains} \quad \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) \left(g_j a_j \frac{\partial}{\partial a_j} + g_j^* a_j^* \frac{\partial}{\partial a_j^*} \right)$$

$$= \ \Re g_j \left(\prod_{k=1}^{n} I_k^{\sigma_k} \right) \left(\frac{1}{a_j^*} \frac{\partial}{\partial a_j} + \frac{1}{a_j} \frac{\partial}{\partial a_j^*} \right) + \Im g_j \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) i \left(a_j \frac{\partial}{\partial a_j} - a_j^* \frac{\partial}{\partial a_j^*} \right)$$

$$= \ 2\Re g_j \left(\prod_{k=1}^{n} I_k^{\sigma_k} \right) \frac{\partial}{\partial I_j} - \Im g_j \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) \frac{\partial}{\partial \varphi_j} \tag{49}$$

If all $\Re g_j = 0$, then we can identify amplitude dependent tuneshifts at degree p by comparison with Eq.(33).

$$\mathbf{N}^{(p)} \ = \ -\sum_{j=1}^{n} \sum_{\underline{\sigma}} \Im g_{j,\underline{\sigma}}^{(p)} \left(I_j^{-1} \prod_{k=1}^{n} I_k^{\sigma_k} \right) \frac{\partial}{\partial \varphi_j} \equiv \sum_{j=1}^{n} \Delta v_j^{(p)}(\underline{I}) \frac{\partial}{\partial \varphi_j}$$

The sum on $\underline{\sigma}$ is carried out only over those exponents satisfying Eq.(43).

Associating this with the Lie generator arising from an autonomous Hamiltonian model is not trivial. We have already seen one piece: it required $\Re g_j = 0$. If some $\Re g_j$ are *significantly* non-zero, it is an indication that the chain of perturbative transformations is breaking down because of accumulation of computational roundoff errors or the existence of a strong resonance that has not been handled correctly. That being assumed, we rewrite

$$\mathbf{N}^{(p)} \ = \ -\sum_{j=1}^{n} \sum_{\underline{\sigma}} \Im g_{j,\underline{\sigma}}^{(p)} \frac{1}{\sigma_j} \frac{\partial}{\partial I_j} \left(\prod_{k=1}^{n} I_k^{\sigma_k} \right) \frac{\partial}{\partial \varphi_j}$$

$$= \ \sum_{j=1}^{n} \frac{\partial}{\partial I_j} \left(-\sum_{\underline{\sigma}} \Im g_{j,\underline{\sigma}}^{(p)} \frac{1}{\sigma_j} \prod_{k=1}^{n} I_k^{\sigma_k} \right) \frac{\partial}{\partial \varphi_j}$$

Clearly we want to identify that final quantity in parentheses as the effective Hamiltonian at degree p,

$$N^{(p)}(\underline{I}) \ = \ \sum_{\underline{\sigma}} N_{\underline{\sigma}}^{(p)} \prod_{k=1}^{n} I_k^{\sigma_k} \ ,$$

but that requires,

$$\textbf{for all } j: \ \textbf{ for all } \underline{\sigma}: \ N_{\underline{\sigma}}^{(p)} \ = \ -\Im g_{j,\underline{\sigma}}^{(p)} / \sigma_j \ , \ \text{independent of } j \ . \tag{50}$$

If that is the case, then we can write, as in Eq.(33),

$$\mathbf{N}^{(p)} \ = \ \sum_{j=1}^{n} \frac{\partial N^{(p)}}{\partial I_j} \frac{\partial}{\partial \varphi_j} \ ,$$

completing the identification, $\Delta v_j^{(p)}(\underline{I}) = \partial N^{(p)}(\underline{I}) / \partial I_j$.

COMMENTS:

41. Proving algebraically that (or when) this will work is beyond the scope of this memo, which at best attempts to lay down a minimal set of the equations implemented in CHEF. In any case, it would be going about the problem backward. If

the original mapping is symplectic, and symplecticity is preserved throughout all orders, then these seeming "conspiracies" among the coefficients are to be expected *unless the perturbation procedure breaks down.* In any case, the normal form series remains valid whether or not the connection with a Hamiltonian can be established. The iterative procedure could even be made to work with non-symplectic mappings.

42. $\Re g_j = 0$ trivially for single thin multipole models, for the kick conditions $\Delta x = 0$ and $\Delta y = 0$ require that $\Delta \underline{a} = -\Delta \underline{a}^*$.

43. The physical consequences of $\Re g_j \neq 0$ are not to be sneezed at. In a continuum interpretation, the coefficient of $\partial/\partial I_j$ is $dI_j/d\theta$. Looking at Eq.(49), a so-called "secular" term is induced when $\Re g_j \neq 0 : d\underline{I}/d\theta$ depends on $\underline{I}$ alone. As has been understood for well over a century, these can lead to divergence and should be avoided when doing perturbation theory.

4.2 Resonance terms

That was the easy part. The treatment of resonance terms – which satisfy Eq.(42) but not Eq.(43) – is more devious. It helps to work backward. Our target is the coefficient of a term in a resonance model Hamiltonian. We postulate that there is a Hamiltonian model, N, like the one in Eq.(34), with terms containing factors $\exp(\pm i\underline{m}^T \cdot \underline{\varphi})$ for which $\sum_{k=1}^{n} m_k v_k = 0 \pmod 1$. We assert further the sufficiency of our basis functions, $\Phi_{\underline{\sigma}}$, so that

$$N \text{ contains resonance terms } N_{\underline{\sigma}}\Phi_{\underline{\sigma}}, \text{ where } m_k = \pm(\sigma_{k+n} - \sigma_k), \text{ for all } k.$$

The resonance condition for the Hamiltonian model can be rewritten.

$$0 \pmod 1 = \sum_{k=1}^{n}(\sigma_k - \sigma_{k+n})v_k = \sum_{k=1}^{n}\sigma_k v_k + \sigma_{k+n}(-v_k) = \sum_{k=1}^{n}\sigma_k v_k + \sigma_{k+n}v_{k+n} = \sum_{k=1}^{2n}\sigma_k v_k \tag{51}$$

The trick is to fix $\underline{m}$ and allow $\underline{\sigma}$ to run over all possibilities.

We are aided and abetted by the geometric nature of vector fields, as expressed in Eq.(32).

$$\begin{aligned}
\mathbf{N} = L_N \text{ contains } N_{\underline{\sigma}}X_{\Phi_{\underline{\sigma}}} &= N_{\underline{\sigma}}\sum_{j=1}^{n}\frac{\partial \Phi_{\underline{\sigma}}}{\partial I_j}\frac{\partial}{\partial \varphi_j} - \frac{\partial \Phi_{\underline{\sigma}}}{\partial \varphi_j}\frac{\partial}{\partial I_j} \\
&= iN_{\underline{\sigma}}\sum_{j=1}^{n}\frac{\partial \Phi_{\underline{\sigma}}}{\partial a_j}\frac{\partial}{\partial a_{j+n}} - \frac{\partial \Phi_{\underline{\sigma}}}{\partial a_{j+n}}\frac{\partial}{\partial a_j} \\
&= iN_{\underline{\sigma}}\sum_{j=1}^{n}\frac{\partial}{\partial a_j}\left(\prod_{k=1}^{2n}a_k^{\sigma_k}\right)\frac{\partial}{\partial a_{j+n}} - \frac{\partial}{\partial a_{j+n}}\left(\prod_{k=1}^{2n}a_k^{\sigma_k}\right)\frac{\partial}{\partial a_j} \\
&= \sum_{j=1}^{n}iN_{\underline{\sigma}}\sigma_j\left(\prod_{k=1}^{2n}a_k^{\sigma_k-\delta_{k,j}}\right)\frac{\partial}{\partial a_{j+n}} - iN_{\underline{\sigma}}\sigma_{j+n}\left(\prod_{k=1}^{2n}a_k^{\sigma_k-\delta_{k,j+n}}\right)\frac{\partial}{\partial a_j}
\end{aligned}$$

We must equate these coefficients with those in the Lie generator of the mapping. Notation again becomes a little awkward. Referring back to Equations (48) and (50), let $g_{j+n,\{\sigma_k-\delta_{k,j}\}}$ be the coefficient of $\prod_k a_k^{\sigma_k-\delta_{k,j}}$ in the component $\mathcal{N}[j+n]$ and let $g_{j,\{\sigma_k-\delta_{k,j+n}\}}$ be the coefficient of $\prod_k a_k^{\sigma_k-\delta_{k,j+n}}$ in $\mathcal{N}[j]$. The postulated relation is,

$$2\pi N_{\underline{\sigma}} = ig_{j,\{\sigma_k-\delta_{k,j+n}\}}/\sigma_{j+n} = -ig_{j+n,\{\sigma_k-\delta_{k,j}\}}/\sigma_j, \text{ for all } j. \tag{52}$$

5 Final comments

We have reached the point of diminishing returns. Rather than continuing in this awkward manner, it would be easier and more meaningful to deal directly with algorithms, their implementations, and case studies of applications, i.e. the subject matter of the third memo.[25] Written here are a few scattered observations before closing.

COMMENTS:

44. Staring back at Equations (41) and (42), it is not obvious that these are indeed the resonance terms singled out previously. That they are is confirmed using Eq.(51).

$$\sum_{k=1}^{2n}((\sigma_k - \delta_{k,j+n}) - \delta_{k,j})\nu_k = \sum_{k=1}^{2n}\sigma_k\nu_k - \sum_{k=1}^{2n}(\delta_{k,j+n} + \delta_{k,j})\nu_k$$
$$= -(\nu_{j+n} + \nu_j) \pmod 1$$
$$= 0 \pmod 1$$

and similarly,
$$\sum_{k=1}^{2n}((\sigma_k - \delta_{k,j}) - \delta_{k,j+n})\nu_k = 0 \pmod 1$$

45. Results analogous to Eq.(52) hold for $\mathcal{T}$, via the components appearing in Eq.(45). If $\mathcal{T}$ can be written as the exponential map generated by a Hamiltonian field, then Eq.(52), properly transposed, must be valid *except that* there is no resonance condition. Thus, Eq.(52) must be applied for *all* exponents, $\underline{\sigma}$. If this works, then the iterative transformations will be symplectic.

46. There are two schools of thought on whether a Hamiltonian interpretation will always work: yes and no. To argue on the positive side, one can appeal to something like the MARYLIE algorithm [5]. It constructs such exponential maps, progressing element by element through a beamline, by using the Baker-Campbell-Hausdorff (BCH) formula[26] to absorb the generators of individual element maps one at a time. On the negative side of the argument, there are statements like the following [13]: "Every element of a *compact* Lie group ... can be obtained by exponentiating some element of the Lie algebra. Often, the theorems that hold for compact groups are no longer valid for noncompact groups. ... For noncompact groups every element may be reached by exponentiating along a small number of straight lines in the Lie algebra."[27] As the symplectic group is not compact, and as we must work with its infinite dimensional representations, this begs the questions, under what conditions (if any) can our techniques fail and are they pathological or of practical concern? How does the issue of convergent versus asymptotic series enter into consideration? One can run but not hide from these questions. I confess to being confused; nevertheless, it is time to move on.

47. This memo is not finished, but I am temporarily dropping it. One missing section has to do with transverse resonances in a storage ring. The resonance formalism written above assumes bounded orbits *in all degrees of freedom*. In accelerator terminology, this means it applies to bunches experiencing possible synchrobetatron resonances. To handle transverse resonances alone we can try (a) taking the limit as cavity voltage and $\delta p/p$ simultaneously approach zero, which may be too cumbersome to consider seriously, or (b) trying to extend the formalism to unbounded orbits (remember the "secular" terms?), or (c) isolating the transverse degrees of freedom by projection and using something like the

---

[25]There might be something to Stephen Wolfram's idea that some physical theories should be - and possibly must be - written as algorithms, not transcribed into analysis.

[26]For a derivation of the BCH formula, see ICD [2, pages 188ff; Problem 3.46, p.196] or practically any textbook on Lie algebras, from which I recommend Gilmore [13].

[27]That other form of exponentiation, "along a small number of straight lines," amounts to an ordered product, like the one shown in Eq.(36).

"co-moving map" approach devised by Irwin and Forest [14], which parallels (almost) exactly the "well known sequence of manipulations" mentioned on page 8, below Eq.(35), which are used in the context of continuous Hamiltonian dynamics [2, pages 104-211]. I shall postpone writing about this until it too is implemented in CHEF's libraries, at which point I may revise this memo or its public successor.

References

[1] Leo Michelotti. Theory and praxis of map analysis in chef; part 1: Linear normal form. Technical Report FN-0826-CD, Fermilab, October 2008.

[2] Leo Michelotti. *Intermediate Classical Dynamics with Applications to Beam Physics*. John Wiley & Sons, Inc., New York, 1995.

[3] Leo Michelotti. Mxyzptlk version 3.1 user's guide: A c++ library for automatic differentiation and differential algebra. Fermi Note FERMILAB-FN-0535-REV, Fermilab, January 31, 1990 (under different title). Revised October, 1995.

[4] Matthew. 23:24.

[5] A. J. Dragt. Lectures on nonlinear orbit dynamics. In *Physics of High Energy Particle Accelerators*, pages 147–313. American Institute of Physics, New York, 1982.

[6] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, Massachusetts, 1982.

[7] A. Deprit. Canonical transformations depending on a small parameter. *Celestial Mechanics*, 1:12–30, 1969.

[8] F. T. Cole. Nonlinear transformations in action-angle variables. Technical report, Fermilab, June 13, 1969. TM-179.

[9] Leo Michelotti. Deprit's algorithm, green's functions and multipole perturbation theory. *Particle Accelerators*, 19:205–210, 1986.

[10] Leo Michelotti. Moser-like transformations using the lie transform. *Particle Accelerators*, 16:233, 1985.

[11] Etienne Forest, Martin Berz, and John Irwin. Normal form methods for complicated periodic systems: A complete solution using differential algebra and lie operators. *Particle Accelerators*, 24:91, 1989.

[12] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley, Reading, Massachusetts, 1959.

[13] Robert Gilmore. *Lie Groups, Lie Algebras, and some of their Applications*. John Wiley & Sons, New York, 1974.

[14] Etienne Forest and John Irwin. Single resonance theory with maps. In W. Scandale and G. Turchetti, editors, *Nonlinear Problems in Future Particle Accelerators*. World Scientific, 1991. Workshop Proceedings. Capri, Italy. April 19-25, 1990.