Optimal separation of mirror and instrumental aberrations

Michael Peck <mpeck1@ix.netcom.com>

December 7, 2010

1 Introduction

This note revisits a topic that arises occasionally on the Yahoo interferometry group, namely how to separate "mirror" from "test stand" aberrations by rotating the "mirror" to two or more orientations. By "mirror" of course we mean the optical system under test, which could be anything capable of forming an image. I will stick to the shorthand "mirror" for this note. The usual context for these discussions is that mechanical pressure from the test stand and gravity distort the mirror under test. Assuming the effects are constant under mirror rotations they can be removed at least partially by testing at multiple orientations. The same procedure can be used to remove instrumental aberrations arising in the interferometer itself. In practice we can't really distinguish instrumental aberrations from test stand induced defects, nor do we need to unless the intent is to calibrate the interferometer. I'm going to use the shorthand "instrumental aberrations" henceforth.

Nothing here is new except perhaps the optimality result proven in section 4. Steve Koehler presented the basic algorithmic approach in message 1941 on the interferometry group, and later in the same thread Michael Koch proposed the same method (with a slightly different optimization function) I will use in section 5 to find the best pair of angles for a 3 rotation set of measurements.

1.1 Zernike polynomials

We don't need to know anything about Zernike polynomials except how they transform under rotations, but I'll review a few key properties here. More extensive discussions are presented in my article on interferometry mathematics, or online at Mathworld.

Zernike polynomials have the functional form:

$$Z_n^m(r,\theta) = \begin{cases} \alpha_n^m R_n^{|m|}(r)\sin(|m|\theta) & m < 0 \\ \alpha_n^m R_n^{|m|}(r)\cos(|m|\theta) & m \geq 0 \end{cases} \tag{1}$$

where $0 \leq r \leq 1$, the indexes n, m satisfy $n \geq 0, -n \leq m \leq n$ with $n - m$ even, and α_n^m is a normalizing factor. The radial functions $R_n^{|m|}(r)$ are n'th order polynomials.

Most applications use a single index scheme to replace the pair n, m. Most optical design and analysis software order by increasing values of $n + |m|$, with aberrations with constant values of $n + |m|$ grouped together and the cosine term first when $m \neq 0$.

Under very general conditions an arbitrary wavefront (or indeed any function defined on the unit disk) can be represented by a weighted sum of Zernike polynomials:

$$W(\boldsymbol{r}) = \sum_{l=0}^{\infty} c_l Z_l(\boldsymbol{r}) \tag{2}$$

In the following all we need to work with are the coefficients c_l, and all we need to know is how they transform under rotations. Note that the axisymmetric aberrations (with $m = 0$) are invariant under rotations. That implies that it is impossible to separate them with this technique. I'm going to say we're "blind" to aberrations that can't be separated by a specific rotation scheme.

Note also that because Zernike polynomials are separable into the product of functions of r and θ we can completely ignore the radial order n.

2 The rotation matrix

If we rotate a wavefront consisting of a pure aberration of azimuthal order m [1] through an angle θ its coefficients transform as

$$\begin{pmatrix} c_1(\theta) \\ c_2(\theta) \end{pmatrix} = \begin{pmatrix} \cos(m\theta) & -\sin(m\theta) \\ \sin(m\theta) & \cos(m\theta) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \tag{3}$$

where c_1, c_2 are the cosine and sine coefficients of the aberration at $\theta = 0$.

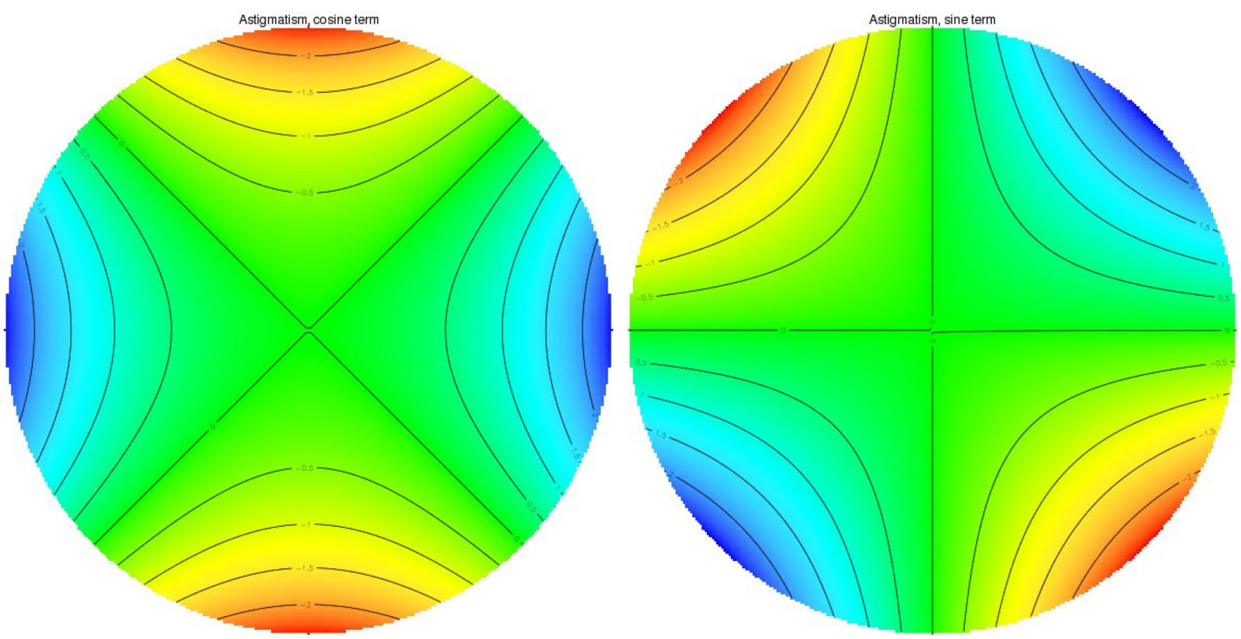

Figure 1: Primary astigmatism

As an example, consider primary astigmatism (figure 1). A 45° counterclockwise rotation of the cosine component produces the sine component. A 90° rotation produces cosine astigmatism with the opposite sign. A 180° rotation produces pure cosine astigmatism again – astigmatism is invariant under a 180°

[1] I will drop the absolute value formalism from here on.

rotation. In general, any aberration of azimuthal order m is invariant under a $360/m$ degree rotation and any integer multiple thereof.

Now suppose we have a component of an aberration that is fixed in our chosen coordinate system and one that rotates with the mirror. The augmented rotation matrix becomes

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & \cos(m\theta) & -\sin(m\theta) \\ 0 & 1 & \sin(m\theta) & \cos(m\theta) \end{pmatrix} \tag{4}$$

Labeling the fixed components t_1, t_2 and the rotating components r_1, r_2 the transformation equation becomes

$$\begin{pmatrix} c_1(\theta) \\ c_2(\theta) \end{pmatrix} = \mathbf{R} \begin{pmatrix} t_1 \\ t_2 \\ r_1 \\ r_2 \end{pmatrix} \tag{5}$$

3 Least squares

This section reviews some features of the linear regression model that will be needed below. This is all standard statistics textbook material – online, Wikipedia offers a decent, if slightly incomplete, treatment.

The most basic form of the linear regression model is conventionally written using matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{6}$$

where $\mathbf{y}$ is a length n vector of measured values of a dependent variable, $\mathbf{X}$ is an n $\times$ p matrix of independent variable values (often called the design matrix), $\boldsymbol{\beta}$ is a length p vector of parameters, and $\boldsymbol{\epsilon}$ is a length n vector of disturbance or error terms.

The ordinary least squares *estimator* for $\boldsymbol{\beta}$ is

$$\hat{\mathbf{b}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y} \tag{7}$$

where T means matrix transpose and $^{-1}$ matrix inverse.

If the random errors are independently and identically distributed with mean 0 and known variance σ^2 the estimates $\hat{\mathbf{b}}$ for $\boldsymbol{\beta}$ have mean $\boldsymbol{\beta}$ (they're unbiased) and covariance matrix

$$\mathbf{V} = \sigma^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} \tag{8}$$

In practice σ is usually unknown, so a sample based estimate s^2 is substituted in equation 8. This won't concern us further here. It's the matrix $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ (or its inverse) that interests us.

Ideally the off diagonal elements of $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ should be as small as possible relative to the diagonal elements. If the independent variables are strongly correlated parameter estimates will be too and will have inflated variances compared to the case where the design matrix $\mathbf{X}$ is orthogonal.

A standard measure of how close a matrix is to diagonal is the condition number, which has several possible definitions. The one I'm going to use is the ratio of the largest to smallest singular value, or actually its inverse, which is bounded by $[0, 1]$. The condition number of a matrix $\mathbf{A}$ is usually denoted by $\kappa(\mathbf{A})$. There seems to be no standard notation for its inverse, so I will use $\nu \equiv 1/\kappa$.

3.1 Application to the matter at hand

Suppose we've measured the mirror at $N \geq 2$ rotation angles $\theta_0, \ldots, \theta_{N-1}$. Without loss of generality we can always assign $\theta_0 = 0$, but for notational convenience I will leave it as a parameter for now.

We create a $2N \times 4$ design matrix by stacking (or interleaving) copies of $\mathbf{R}$ (equation 4):

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & \cos(m\theta_0) & -\sin(m\theta_0) \\ 0 & 1 & \sin(m\theta_0) & \cos(m\theta_0) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cos(m\theta_{N-1}) & -\sin(m\theta_{N-1}) \\ 0 & 1 & \sin(m\theta_{N-1}) & \cos(m\theta_{N-1}) \end{pmatrix} \tag{9}$$

Calculating $\mathbf{R^TR}$ is a relatively straightforward pencil and paper exercise:

$$\mathbf{R^TR} = \begin{pmatrix} N & 0 & \sum_{n=0}^{N-1}\cos(m\theta_n) & -\sum_{n=0}^{N-1}\sin(m\theta_n) \\ 0 & N & \sum_{n=0}^{N-1}\sin(m\theta_n) & \sum_{n=0}^{N-1}\cos(m\theta_n) \\ \sum_{n=0}^{N-1}\cos(m\theta_n) & \sum_{n=0}^{N-1}\sin(m\theta_n) & N & 0 \\ -\sum_{n=0}^{N-1}\sin(m\theta_n) & \sum_{n=0}^{N-1}\cos(m\theta_n) & 0 & N \end{pmatrix} \tag{10}$$

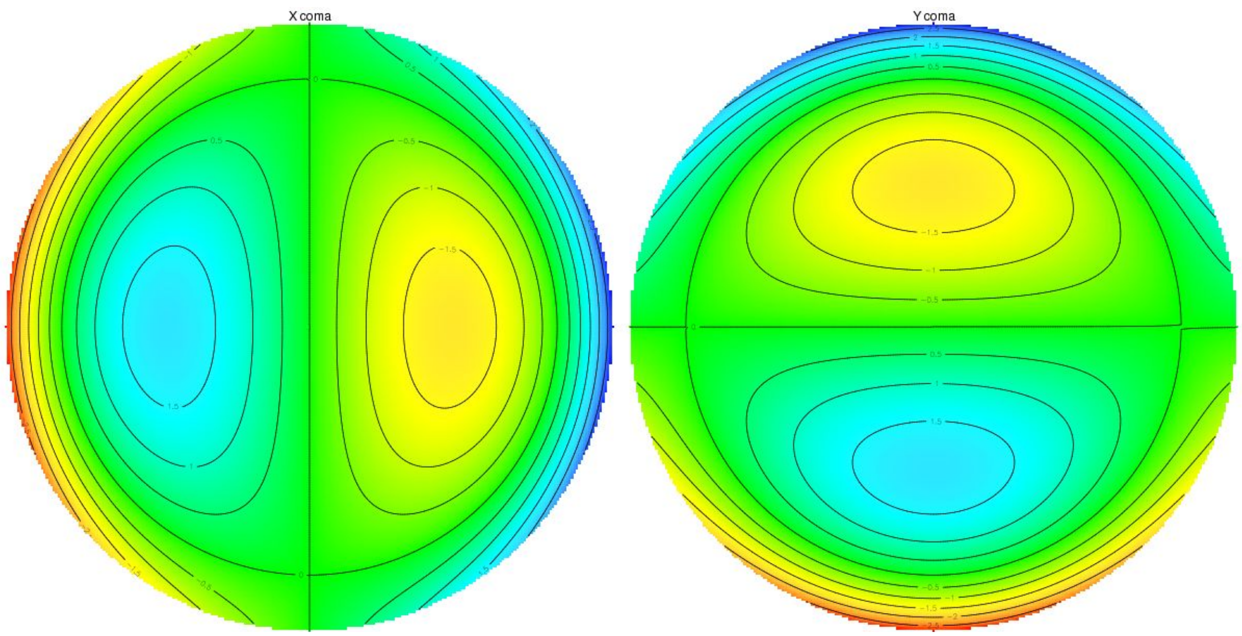

Figure 2: Coma

As an example, for $m = 1$ (coma and higher order versions of coma) $180°$ is clearly the optimum rotation angle for a single pair of measurements. This is more or less visually obvious in figure 2: a $180°$ rotation of X or Y coma produces X or Y coma with the opposite sign.

Since $\cos(\pi) = -\cos(0) = -1$ and $\sin(\pi) = \sin(0) = 0$ $\mathbf{R^TR}$ reduces to

$$
\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}
$$

and since the matrix is diagonal its inverse condition number is $\nu = 1$.

Suppose we use a not so favorable rotation, say $90°$. Plugging into equation 10 we get

$$
\mathbf{R^T R} = \begin{pmatrix} 2 & 0 & 1 & -1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ -1 & 1 & 0 & 2 \end{pmatrix}
$$

and its inverse is

$$
(\mathbf{R^T R})^{-1} = \begin{pmatrix} 1 & 0 & -1/2 & 1/2 \\ 0 & 1 & -1/2 & -1/2 \\ -1/2 & -1/2 & 1 & 0 \\ 1/2 & -1/2 & 0 & 1 \end{pmatrix}
$$

From equation 8 we see that this case will have twice the variance in coefficient estimates, assuming equal measurement errors, and significantly less clean separation of mirror and instrumental coma. The inverse condition number for this matrix by the way is $\nu \simeq 0.17$.

As a final remark, equation 10 allows us to infer immediately the optimal rotation angle for any single aberration. It is π/m radians, or $180/m$ degrees. This is a well known result.

4 The main result

Proposition: Suppose we make measurements at N equally spaced angles $\theta_n = 0, 2\pi/N, \ldots, 2\pi(N-1)/N$. This is optimal in the sense discussed in the previous section for all m **except** $m = 0, N, 2N, 3N, ...$, for which we are completely unable to separate mirror from instrumental aberrations.

Proof: What I show is that the matrix $\mathbf{R^T R}$ in equation 10 is either diagonal or singular.

Proving the except condition is almost trivially simple. If $m = N$, $\cos(2\pi nm/N) = \cos(2\pi n) = 1 \ \forall n$, similarly $\sin(\cdot) = 0 \ \forall \ n$, and the matrix $\mathbf{R^T R}$ reduces to

$$
\begin{pmatrix} N & 0 & N & 0 \\ 0 & N & 0 & N \\ N & 0 & N & 0 \\ 0 & N & 0 & N \end{pmatrix}
$$

Which is clearly of rank 2 and hence singular. The same argument applies for m any integer multiple of N.

To prove the first part, we need to show that $\sum \cos(\cdot) = \sum \sin(\cdot) = 0$. It turns out to be easier to work with the complex exponential $\exp(\mathrm{i}\theta) \equiv \cos(\theta) + \mathrm{i}\sin(\theta)$, which will allow us to prove both conditions in one pass. We make use of a standard result of elementary analysis[2]:

$$\sum_{n=0}^{N-1} r^n = \frac{1 - r^N}{1 - r} \quad \text{for } r \notin \{0, 1\} \tag{11}$$

r can be real or complex. Plugging $r = \exp(\mathrm{i}2\pi m/N)$ into equation 11

$$\sum_{n=0}^{N-1} \exp(\mathrm{i}2\pi nm/N) = \frac{1 - \exp(\mathrm{i}2\pi m)}{1 - \exp(\mathrm{i}2\pi m/N)} = 0 \tag{12}$$

since $\exp(\mathrm{i}2\pi m) = 1 \ \forall \ m$. The only other conditions we need to check are that $r = \exp(\mathrm{i}2\pi m/N)$ is neither 0 nor 1. We can immediately exclude 0, because $|\exp(\mathrm{i}\theta)| = 1$ always. Also, $\exp(\mathrm{i}2\pi m/N) = 1$ only for m/N integer, which is the condition we excluded. Therefore $\mathbf{R^T R}$ is diagonal:

$$\begin{pmatrix} N & 0 & 0 & 0 \\ 0 & N & 0 & 0 \\ 0 & 0 & N & 0 \\ 0 & 0 & 0 & N \end{pmatrix}$$

How many sets of measurements are enough? This is a judgement call obviously, but clearly $N = 2$ or 3 are inadequate since $N = 2$ is blind to all even azimuthal order aberrations including primary astigmatism, while $N = 3$ is blind to trefoil and its cousins.

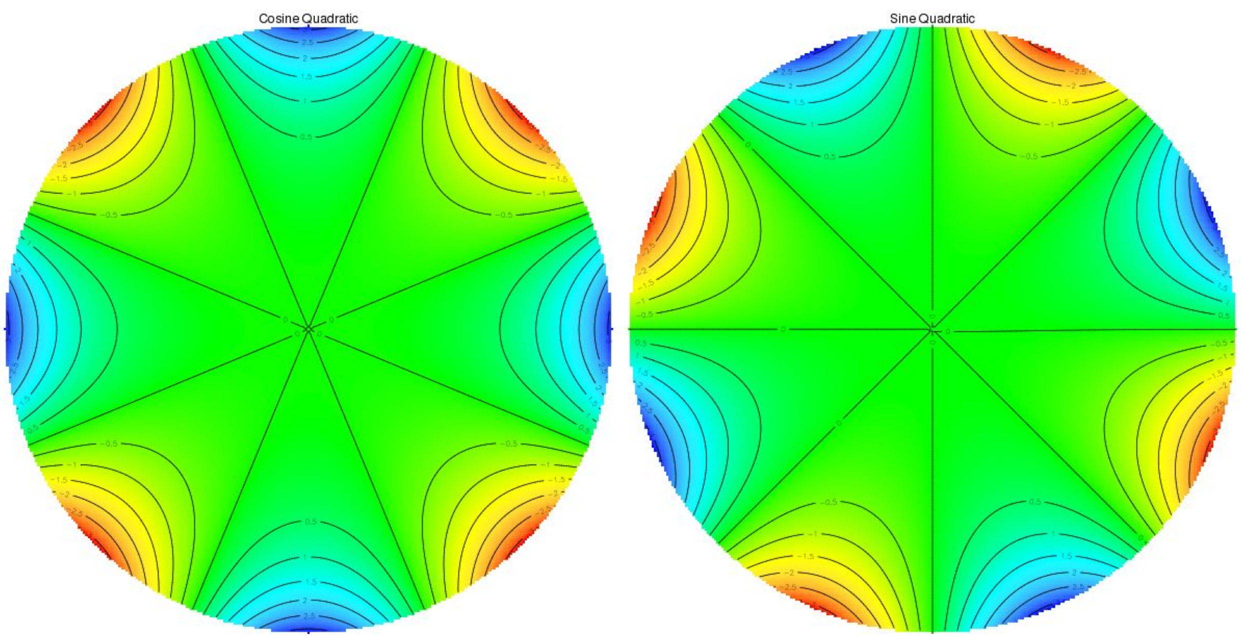

Figure 3: Quadratic

At $N = 4$ the lowest order aberration we're blind to is what is commonly called "quadratic" (Figure 3). An aberration like this *might* be nontrivial in a mirror flexing under its own weight.

[2] I'll derive this in the appendix

5 Max-min strategies

Again, there is nothing fundamentally new here. Most of the ideas were developed by Steve Koehler and Michael Koch in the Interferometry group thread starting in message 1941, with Michael Koch proposing what is essentially the max-min strategy investigated here in message 1999, with a slightly different optimization target.

The idea is that if we are interested in aberrations up to order M with N sets of measurements a reasonable conservative strategy to pick a set of rotation angles is to maximize the worst case condition number. Formally, we want to solve

$$\max_{\theta_1,\ldots,\theta_{N-1}} \min_{m=1,\ldots,M} \nu(m, \boldsymbol{\theta}) \tag{13}$$

where $\nu(\cdot)$ is the inverse condition number introduced in section 3, with the dependence on m and $\boldsymbol{\theta}$ added to make it explicit. The angles θ_n might or might not be required to be evenly spaced. A few numerical examples will suffice.

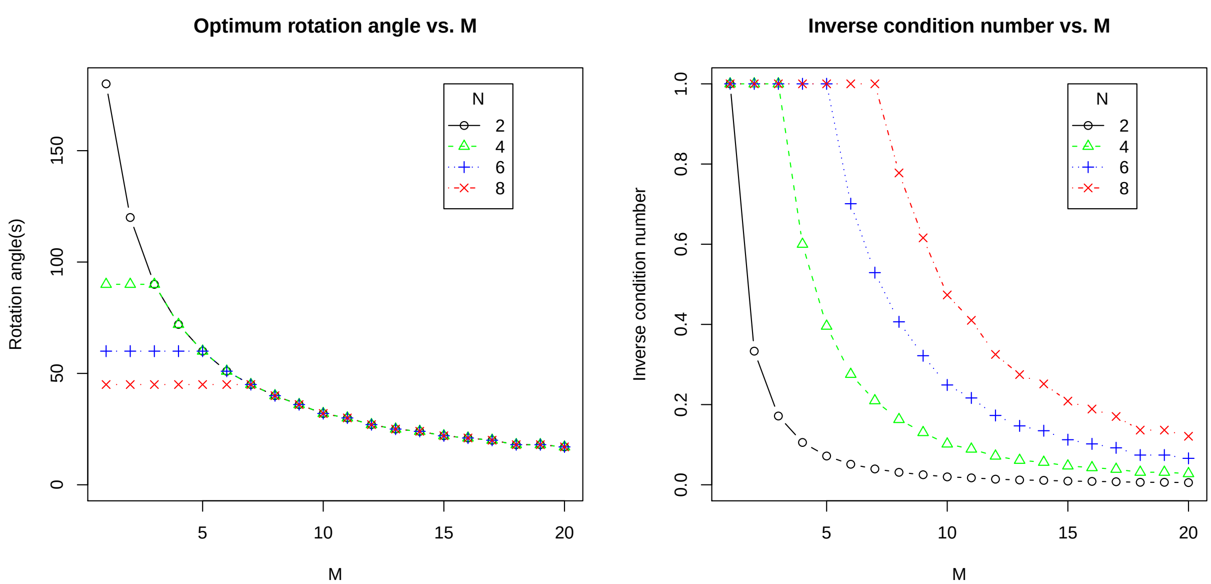

Figure 4: Optimum rotation angles and max-min condition numbers for N measurements vs. M

First, suppose we make N sets of equally spaced measurements and we want to fit up to order M. Figure 4 shows the results for $N = 2, 4, 6, 8$ and M from 1 to 20 (which is much higher order than we should normally care about). A couple of results displayed in these graphs are expected. First, more measurements are always better as seen in the right hand graph. Second, for $M < N$ it's always optimal to select $\Delta\theta = 360°/N$, which is easily inferred from the previous section. What was less expected to me at least is that all of the plotted examples follow the same curve downward for high M, and that curve is, within the rounding errors of the search, $360°/(M + 1)$. That of course is the angle we would pick if we made $N = M + 1$ measurements to fit up to order M.

As a second, and final, example in message 1999 Michael Koch suggested that it might be preferable to make measurements at *unequal* angles between sets, and made some numerical estimates for N=3 and varying M. Figure 5 shows my attempt to replicate this graph and this one by Steve Koehler (both links were live at the time this was written) for a few values of M. These graphs map the value of ν, shown as gray scale levels,

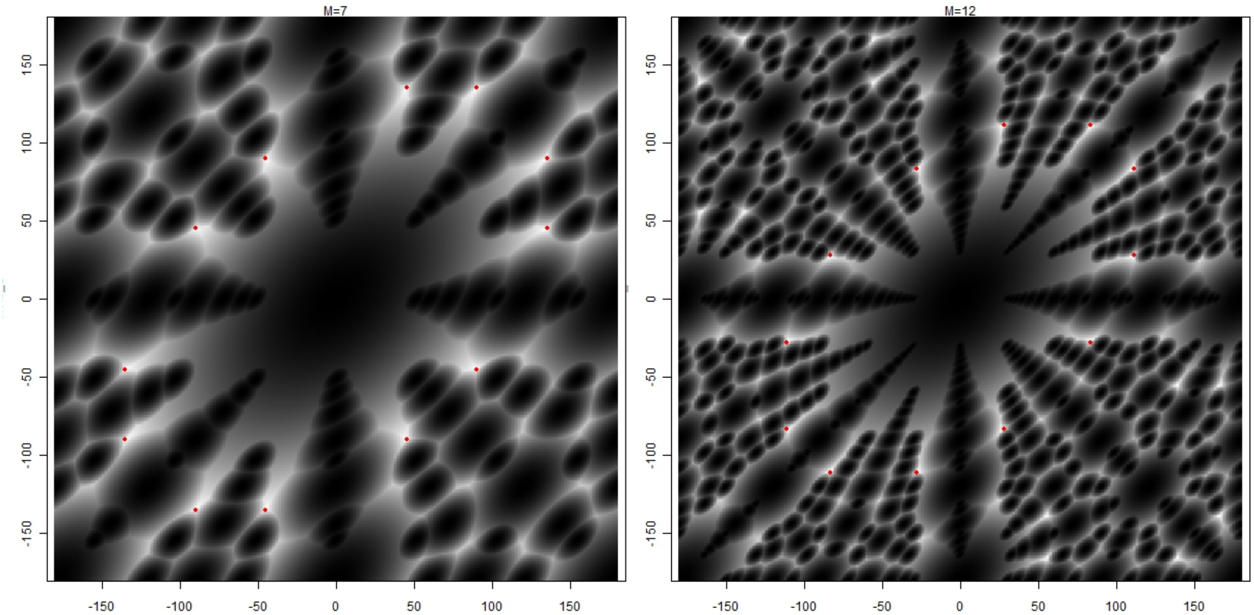

Figure 5: Condition number as a function of θ_1, θ_2

against θ_1 and θ_2, with optima marked with red dots. Although the details differ slightly, the maps show similar levels of complexity, and for the cases I've checked one of the optima is always within a degree or so of the ones calculated by Michael Koch.

Using unequal rotation angles does indeed improve results. A graph of the inverse condition number against M generally falls between the N=4 and N=6 curves in figure 4. It would be interesting to extend this analysis to higher N. Unfortunately the fascinating, even beautiful, seemingly fractal landscape of ridges and valleys seen in these maps makes it likely that exhaustive search must be used to find optima.

A Derivation of equation 11

The derivation makes use of a simple trick.

First note that if r is 0 or 1 the sum in equation 11 is trivially either 0 or N, so we must exclude those two values. For $r \notin \{0, 1\}$,

$$\begin{aligned}
r \sum_{n=0}^{N-1} r^n &= \sum_{n=1}^{N} r^n \\
&= \sum_{n=0}^{N-1} r^n - 1 + r^N \\
(1-r) \sum_{n=0}^{N-1} r^n &= 1 - r^N \\
\sum_{n=0}^{N-1} r^n &= (1 - r^N)/(1-r)
\end{aligned}$$