Towards a Geometrical Understanding of Undermodelled CMA

Phil Bert Schniter

October 9, 1997

Abstract: This report describes some preliminary results aimed at understanding the topology of the undermodelled CMA 2-2 cost surface in the absence of noise and common subchannel roots.

1 Introduction and Notation

One could argue that principle difficulty in analyzing the topology of CMA in the presence of undermodelling is due to the fact that not all channel-eqalizer responses are "reachable". The existing results on the topology of CMA depend heavily on the reachability of particular locations in system space. The following analysis reformulates the historic system-space view in terms of the reachable system subspace.

We denote the channel impulse response vector by $\mathbf{h}$, the equalizer impulse response vector by $\mathbf{f}$, and the system impulse response vector by $\mathbf{q}$. The channel convolution matrix $\mathbf{H}$ relates the equalizer and system responses via $\mathbf{q} = \mathbf{H}\mathbf{f}$. The size and structure of the channel convolution matrix is determined by the sizes of $\mathbf{h}$ and $\mathbf{q}$ and the degree of fractional sampling, P. Presently this does not enter into our analysis, as we rely only on the dimension of $\mathbf{H}$ ($N_q \times N_f$). It should be noted, however, that as a convolution matrix, $\mathbf{H}$ must have a Sylvester form governed by the value of P.

The case of channel undermodelling implies that $N_q > N_f$. For simplicity, we assume that $\mathbf{H}$ is full column rank (though *not* full row rank) and real-valued.

2 Channel Decomposition

The principle tool used in this study of undermodelled CMA is the singular value decomposition (SVD). We apply this to the channel convolution matrix as follows.

$$\mathbf{H} \;=\; \mathbf{U}\mathbf{S}\mathbf{V}^t. \tag{1}$$

For full column rank $\mathbf{H}$, the SVD specifies that $\mathbf{U}$ is a $N_q \times N_q$ orthogonal matrix composed of the left singular vectors, $\mathbf{V}$ is a $N_f \times N_f$ orthogonal matrix composed of the right singular vectors, and $\mathbf{S}$ is a $N_q \times N_f$ matrix with non-zero elements (the "singular values") only along the uppermost diagonal.

Applying the channel decomposition to the system response, we have $\mathbf{q} = \mathbf{U}\mathbf{S}\mathbf{V}^t\mathbf{f}$. Using the orthogonal property of $\mathbf{U}$,

$$\mathbf{p} \;\stackrel{\triangle}{=}\; \mathbf{U}^t\mathbf{q} \;=\; \mathbf{S}\mathbf{V}^t\mathbf{f}. \tag{2}$$

The *transformed system response* $\mathbf{p}$ has an important property: for any choice of equalizer parameters $\mathbf{f}$, the last $N_q - N_f$ entries in $\mathbf{p}$ are zero, as implied by the structure of $\mathbf{S}$. Hence we can partition our quantities to take advantage of this structure.

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{p}_1 \\ 0 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{S}_1 \\ 0 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix}.$$

Both $\mathbf{U}_1$ and $\mathbf{S}_1$ are square matrices of size $N_f \times N_f$. We refer to the $N_f \times 1$ vector $\mathbf{p}_1$ as the *reachable transformed system response*. In terms of the partioned quantities, we have the fundamental relationships

$$\begin{aligned} \mathbf{H} &= \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}^t, & (3) \\ \mathbf{p}_1 &= \mathbf{S}_1 \mathbf{V}^t \mathbf{f}, & (4) \\ \mathbf{q} &= \mathbf{U}_1 \mathbf{p}_1. & (5) \end{aligned}$$

With the invertibility of $\mathbf{S}_1 \mathbf{V}^t$, we establish that $\mathbf{p}_1$ and $\mathbf{f}$ are isomorphic.

3 CMA Cost Function

In the absence of noise and common subchannel roots, the real-valued CMA 2-2 cost function can be specified in terms of its system response $\mathbf{q}$ [1] as

$$J_{\mathrm{CMA}}(\mathbf{q}) = (\kappa_s - 3)\mathbf{q}^t \mathrm{diag}(\mathbf{q}\mathbf{q}^t)\mathbf{q} + 3\|\mathbf{q}\|_2^4 - 2\kappa_s \|\mathbf{q}\|_2^2 + \kappa_s^2 \quad (6)$$

where the $\mathrm{diag}(\cdot)$ preserves the diagonal elements of its matrix argument and zeros the rest. Equation (5) implies that, in terms of the reachable $\mathbf{p}_1$,

$$J_{\mathrm{CMA}}(\mathbf{p}_1) = (\kappa_s - 3)\mathbf{p}_1^t \mathbf{U}_1^t \mathrm{diag}(\mathbf{U}_1 \mathbf{p}_1 \mathbf{p}_1^t \mathbf{U}_1^t)\mathbf{U}_1 \mathbf{p}_1 + 3\|\mathbf{p}_1\|_2^4 - 2\kappa_s \|\mathbf{p}_1\|_2^2 + \kappa_s^2 \quad (7)$$

where we have employed the fact that $\mathbf{U}_1^t \mathbf{U}_1 = \mathbf{I}_{N_f}$.

4 Gradient in Equalizer Space

In this section we derive a compact expression for the gradient of the undermodelled CMA cost function in equalizer space, $\nabla_{\mathbf{f}} J_{\mathrm{CMA}}$. We find the chain rule convenient

$$\begin{aligned} \nabla_{\mathbf{f}} J_{\mathrm{CMA}} &= \nabla_{\mathbf{f}} \mathbf{p}_1 \; \nabla_{\mathbf{p}_1} J_{\mathrm{CMA}} \\ &= \mathbf{V}\mathbf{S}_1 \, \nabla_{\mathbf{p}_1} J_{\mathrm{CMA}} \end{aligned}$$

since $\nabla_{\mathbf{p}_1} J_{\mathrm{CMA}}$ can be calculated using the identities

$$\begin{aligned} \nabla_{\mathbf{p}_1} \left(\mathbf{p}_1^t \mathbf{U}_1^t \mathrm{diag}(\mathbf{U}_1 \mathbf{p}_1 \mathbf{p}_1^t \mathbf{U}_1^t)\mathbf{U}_1 \mathbf{p}_1 \right) &= 4\mathbf{U}_1^t \mathrm{diag}\{(\mathbf{w}_1^t \mathbf{p}_1)^2, \ldots, (\mathbf{w}_N^t \mathbf{p}_1)^2\}\mathbf{U}_1 \mathbf{p}_1 & (8) \\ \nabla_{\mathbf{p}_1} \|\mathbf{p}_1\|_2^4 &= 4\|\mathbf{p}_1\|_2^2 \mathbf{p}_1 & (9) \\ \nabla_{\mathbf{p}_1} \|\mathbf{p}_1\|_2^2 &= 2\mathbf{p}_1 & (10) \end{aligned}$$

where $\mathbf{w}_n^t$ specifies the n^{th} row[1] of $\mathbf{U}_1$ and where the operation $\mathrm{diag}\{\cdot, \ldots, \cdot\}$ returns a diagonal matrix constructed from its scalar arguments.

[1] Recall that, unlike the columns, the rows of $\mathbf{U}_1$ do *not* form an orthonormal set.

Combining the gradient identities (8)-(10) and substituting into (8), we arrive at the compact expression

$$\nabla_{\mathbf{f}} J_{\text{CMA}} \quad = \quad \mathbf{H}^t \mathbf{D}(\mathbf{p}_1)\mathbf{U}_1\mathbf{p}_1 \tag{11}$$

where $\mathbf{D}(\mathbf{p}_1)$ is a diagonal matrix specified by the elements:

$$[\mathbf{D}(\mathbf{p}_1)]_{n,n} \quad = \quad 4\big((\kappa_s - 3)(\mathbf{w}_n^t \mathbf{p}_1)^2 + 3\|\mathbf{p}_1\|_2^2 - \kappa_s\big) \tag{12}$$

The result in (11) is a generalization of the following gradient expression, derived in the context of invertible $\mathbf{H}$ [1].

$$\nabla_{\mathbf{f}} J_{\text{CMA}} \quad = \quad 4\mathbf{H}^t \text{diag}\{\ldots, (\kappa_s - 3)q_n^2 + 3\|\mathbf{q}\|_2^2 - \kappa_s, \ldots\}\mathbf{q} \tag{13}$$

It is straightforward to verify that (11) and (12) lead to (13) when $\mathbf{U}_1$ is an orthogonal matrix.

The equalizer-space gradient may be used to catalog the CMA stationary points, i.e. the points where $\nabla_{\mathbf{f}} J_{\text{CMA}} = 0$. This is straightforward in the case that $\mathbf{H}$ is full row rank, where the requirement that $\mathbf{D}(\mathbf{p}_1)\mathbf{U}_1\mathbf{p}_1 = 0$ follows from (11). From (13) it can be seen that this zero-gradient requirement is satisfied when

$$\big((\kappa_s - 3)q_n^2 + 3\|\mathbf{q}\|_2^2 - \kappa_s\big)q_n = 0 \text{ for all } n$$

or, equivalently, when the values q_n are chosen from the set $\left\{0, \pm\sqrt{\frac{3\|\mathbf{q}\|_2^2 - \kappa_s}{3 - \kappa_s}}\right\}$. Here we have used the facts that $\mathbf{w}_n^t \mathbf{p}_1 = q_n$ and $\|\mathbf{p}_1\| = \|\mathbf{q}\|$.

When $\mathbf{H}$ is not full row rank, as is the case with channel undermodelling, the situation is more complicated. From (11) and (3), the necessary and sufficient zero-gradient condition becomes

$$\mathbf{U}_1^t \mathbf{D}(\mathbf{p}_1)\mathbf{U}_1\mathbf{p}_1 = 0.$$

Thus, the zero-gradient condition becomes

$$\mathbf{D}(\mathbf{p}_1)\mathbf{U}_1\mathbf{p}_1 \in \text{span}\{\mathbf{U}_2\} \tag{14}$$

where $\text{span}\{\mathbf{U}_2\}$ denotes the space spanned by the columns of $\mathbf{U}_2$.

Equation (14) can be put in a form which may be more intuitive. Using the facts that $\mathbf{q} = \mathbf{U}_1\mathbf{p}_1$ and $q_n = \mathbf{w}_n^t\mathbf{p}_1$, and the notation $\hat{\mathbf{q}} = \mathbf{q}/\|\mathbf{q}\|$ for the unit normalized system response, the gradient is zeroed when

$$(\hat{q}_1^3 + a\hat{q}_1, \ldots, \hat{q}_{N_f}^3 + a\hat{q}_{N_f}) \in \text{span}\{\mathbf{U}_2\} \quad \text{for} \quad \hat{\mathbf{q}} \in \text{span}\{\mathbf{U}_1\} \tag{15}$$

where a is the $\|\mathbf{q}\|$-dependent scalar

$$a \quad = \quad -\frac{3\|\mathbf{q}\|_2^2 - \kappa_s}{\|\mathbf{q}\|_2^2(3 - \kappa_s)}. \tag{16}$$

Unfortunately, we have not yet been able to count the number of stationary points based on (14) and (15).

5 Hessian in Equalizer Space

The Hessian has not been derived yet.

6 Geometrical Interpretation

In this section we give a geometrical interpretation of the results in the previous sections. We will begin with the case where $N_q = 2$ and $N_f = 1$ and follow with the case where $N_q = 3$ and $N_f = 2$.

Throughout this report we take advantage of the special structure of the noiseless CMA 2-2 cost function in N_q-dimensional system space. Specifically, the zero forcing (ZF) system responses[2] $\{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_{N_q}\}$ are known to minimize the CMA cost function, assuming that the source kurtosis κ_s is sub-gaussian. (In the absence of noise, the ZF system responses are optimal with respect to any reasonable cost criterion since they result in perfect symbol recovery.) In the $N_q = 2$ case, these ZF $\mathbf{q}$ vectors form a "unit cross" aligned with the coordinate axes in system space (see Fig. 1). In the $N_q = 3$ case, the ZF $\mathbf{q}$ vectors resemble the minature metal

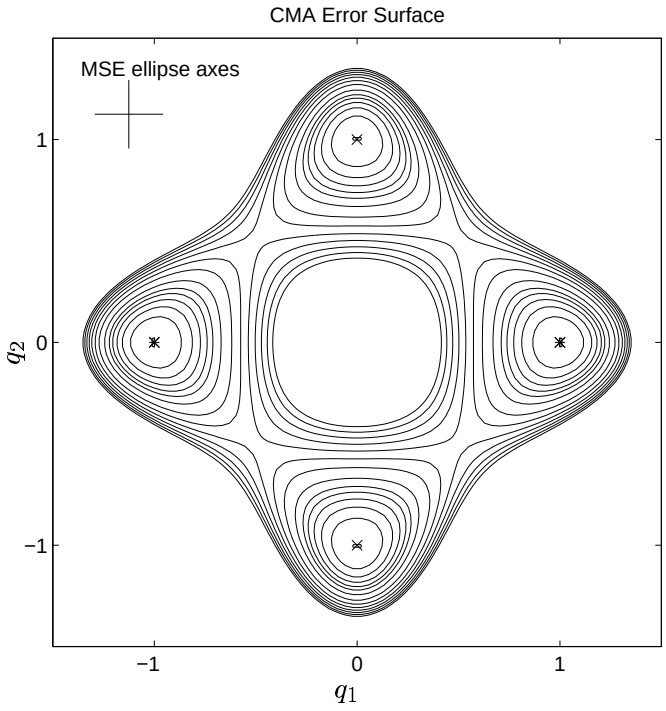

Figure 1: J_{cm} contours for $kappa_s = 1$ (BPSK) and no noise, in system ($\mathbf{q}$) space.

objects used to play the game "jacks". Such visualization will soon prove to be insightful.

The SVD also has geometrical properties which we make use of. First, multiplication by an N-dimensional orthogonal matrix results in a pure N-dimensional rotation, i.e. a rotation devoid of any dilation. In constrast, multiplication by an N-dimensional diagonal matrix results in a dilation along each coordinate axis (without any rotation). Putting these facts together, we can view the SVD as decomposing a $N_q \times N_f$ (where $N_q > N_f$) matrix into the following four consecutive operations:

1. rotation in N_f-dimensional space (via $\mathbf{V}^t$).
2. dilation in N_f-dimensional space (via $\mathbf{S}_1$).

[2] We denote the unit vector with non-zero element in the n^{th} position by $\mathbf{e}_n$.

3. the embedding of the N_f-dimensional subspace in the N_q-dimensional space (via $\mathbf{S}_2$). (Specifically, the low dimensional subspace constitues a hyperplane spanned by the first N_f coordinate axes of the higher dimensional space.)

4. rotation in N_q-dimensional space (via $\mathbf{U}$).

Note that $\mathbf{U}^t$ corresponds to the "opposite" rotation as $\mathbf{U}$, since $\mathbf{U}^t\mathbf{U} = \mathbf{U}\mathbf{U}^t = \mathbf{I}$.

6.1 Two-dimensional Example

Here we consider the case where $N_q = 2$ and $N_f = 1$, or in other words when the channel convolution matrix is of the form $\mathbf{H} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$. The reachable system subspace consists of a line in the plane of system parameters $\mathbf{q}$, as in Fig. 2(a). Transforming the elements in $\mathbf{q}$-space by $\mathbf{U}^t$ rotates the reachable subspace to the $\mathbf{e}_1$ coordinate axis, as depicted by the $\mathbf{p}$-space plot Fig. 2(b).

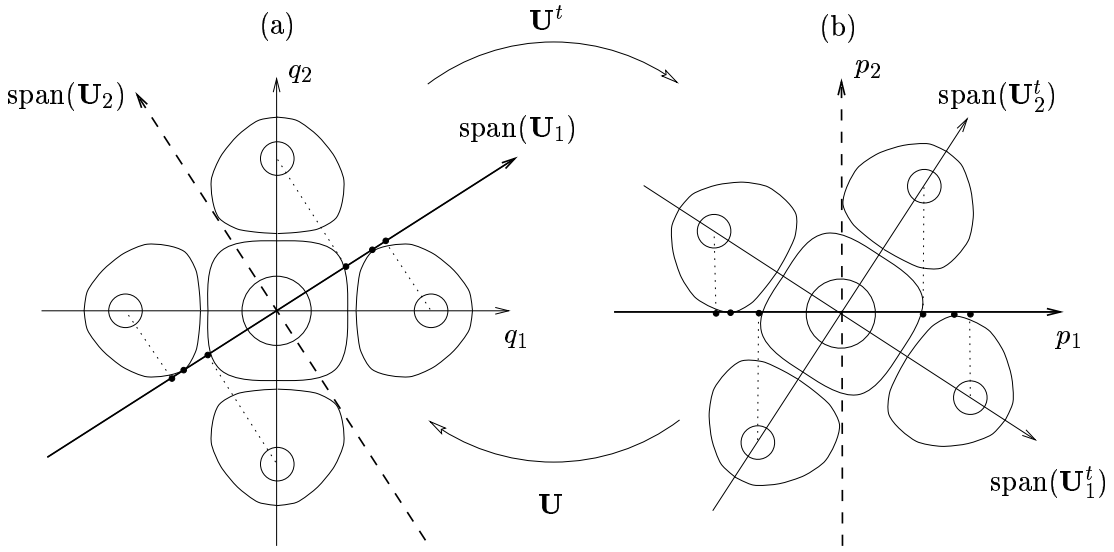

Figure 2: Drawing of CMA 2-2 contours in (a) system-space $\mathbf{q}$ and (b) transformed system space $\mathbf{p}$. CMA and projected MMSE solutions are denoted by dots on the reachable subspace.

The reachable system responses minimizing CM cost occur at the tangent intersections of the CM contours and the reachable subspace. We can use the same argument to determine the MMSE solutions, using the fact that the MSE cost contours (corresponding to a given system delay/polarity) take the form of circles centered at the respective ZF solution in $\mathbf{q}$- or $\mathbf{p}$-space. Thus, the reachable MMSE reachable responses are the projections of the respective ZF responses onto the reachable subspace. These projections are illustrated in Fig. 2. It follows that the MSE cost associated with a particular solution can be measured by the distance of the respective ZF solution from the reachable subspace.

The previous discussion enables us to make an important observation about the discrepancy between the MMSE and CMA solutions: it is caused by the fact that the CM cost contours are not circular in a Euclidean sense (see Fig. 1). Note, however, that the CM cost contours become more circular as they approach the ZF solutions. This can be used to explain a claim made in [2]:

The CM minima corresponding to system delays which give better MSE performance stay closer to their corresponding Wiener solutions than the CM minima corresponding to system delays which give poorer MSE performance.

6.2 Three-dimensional Examples

The three-dimensional examples in the following subsections extend the intuition gained from the two-dimensional example of Section 6.1. The CM contours (see for example Fig. 3) can be considered a 2-dimensional slice of a multi-layered 3-dimensional object, the layers of which form surfaces of uniform CM cost. Likewise, the MMSE solutions are the projections of the $\mathbf{q}$-space ZF solutions onto the $\mathbf{p}_1$ plane.

Here we consider $N_q = 3$ and $N_f = 2$. For a $T/2$-spaced equalizer, these dimensions specify a channel convolution matrix of the form

$$\mathbf{H} = \begin{pmatrix} h_2 & h_1 \\ h_4 & h_3 \\ h_6 & h_5 \end{pmatrix}.$$

When constructing examples, the lack of structure in $\mathbf{H}$ implies that we are free to choose the matrices $\mathbf{U}, \mathbf{S}_1$, and $\mathbf{V}$. Appendix A specifies a method of constructing a rotation matrix in terms of rotation angles $\{\phi_1, \phi_2, \ldots, \phi_N\}$ between pairs of coordinate axes $\{(\mathbf{e}_1, \mathbf{e}_2), (\mathbf{e}_2, \mathbf{e}_3), \ldots, (\mathbf{e}_N, \mathbf{e}_1)\}$, respectively.

Henceforth, it helps to visualize a minature metal "jack" as the shape representing the six $\mathbf{q}$-space ZF responses (see Fig. 4(a)). We are interested in choosing rotation matrices $\mathbf{U}^t$ that spherically rotate this jack until it forms interesting relationships with the horizontal plane (i.e. the plane spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$), since we know this plane corresponds to the transformed reachable system subspace. Figure 4(b) shows an example of one such rotation.

Having already established the fact that $\mathbf{p}_1$ and $\mathbf{f}$ are isomorphic, we choose (without loss of generality) $\mathbf{S}_1\mathbf{V}^t = \mathbf{I}_{N_f}$ in the following examples, allowing the simplification $\mathbf{f} = \mathbf{p}_1 = \mathbf{U}_1^t\mathbf{h}$.

6.2.1 A "Typical" Scenario

We consider a channel constructed using the set of arbitrarily-chosen rotation angles $\{\pi/3, \pi/8, \pi/3\}$ to be "typical" in terms of its effect on the CM cost surface in $\mathbf{f}$-space. Figure 3 plots these CM cost contours with the markers $*$ and $\times$ specifying the locations of global and local MMSE solutions, respectively. Note the presence of two pairs of distinct CM minima in which the deeper CM minima have MMSE solutions in closer proximity.

6.2.2 Undermodelled yet Perfect Equalization

Certain "trivial" rotations can be used to explain the possibility of perfect symbol recovery in the presence of channel undermodelling. The existence of a perfectly equalizing solution requires that at least one pair of ZF system responses lies in the reachable subspace. For example, the structure of Fig. 4(a) is preserved by rotation angles ϕ_n constrained to a multiples of $\pi/4$.

To be more general, the only requirement for perfect symbol recovery is that one of the rows of $\mathbf{U}$ lies in the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. This can be translated into a decomposition-based condition on the channel convolution matrix or may be used to specify a class of rotation angles which can be used to create $\mathbf{U}_1$ (and hence $\mathbf{H}$). In the 3×3 case, Appendix A specifies the following

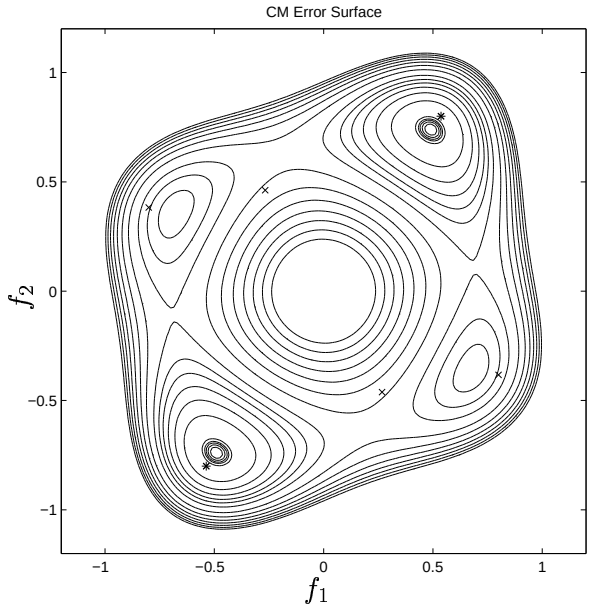

Figure 3: CM cost contours resulting from rotation angles $\{\pi/3, \pi/8, \pi/3\}$ in $(\mathbf{p}_1 =)\mathbf{f}$-space.

relationship between $\mathbf{U}$ and $\{\phi_1, \phi_2, \phi_3\}$:

$$\mathbf{U} = \begin{pmatrix} \cos\phi_3 \cos\phi_1 + \sin\phi_3 \sin\phi_2 \sin\phi_1 & -\cos\phi_3 \sin\phi_1 + \sin\phi_3 \sin\phi_2 \cos\phi_1 & \sin\phi_3 \cos\phi_2 \\ \cos\phi_2 \sin\phi_1 & \cos\phi_2 \cos\phi_1 & -\sin\phi_2 \\ -\sin\phi_3 \cos\phi_1 + \cos\phi_3 \sin\phi_2 \sin\phi_1 & \sin\phi_3 \sin\phi_1 + \cos\phi_3 \sin\phi_2 \cos\phi_1 & \cos\phi_3 \cos\phi_2 \end{pmatrix}$$

6.2.3 Loss of Distinct Minima

Perhaps the most interesting channel convolution matrices can be constructed from matrices that rotate the ZF solutions onto a plane parallel to the reachable subspace (see Fig. 4(b)). We say such channels belong to the *MMSE maximizing class*: $\{\arg\max_{\mathbf{h}} \min_{\delta,\mathbf{f}} J_{\mathrm{MSE}}\}$.

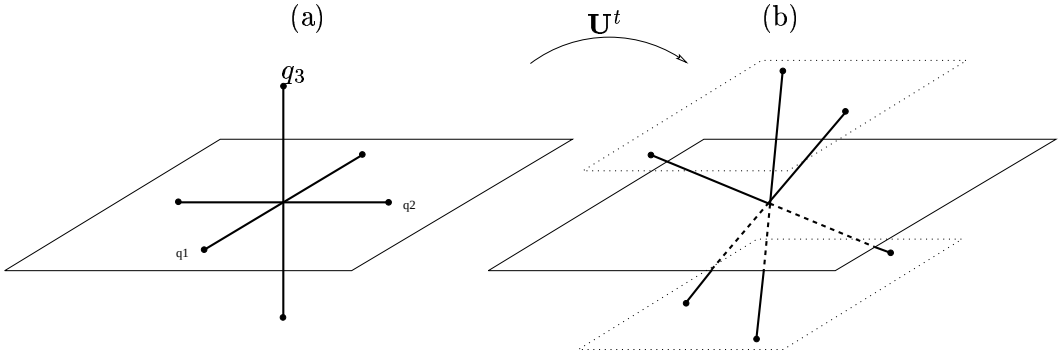

Figure 4: (a) ZF equalizers in $\mathbf{q}$-space, (b) ZF equalizers in $\mathbf{p}$-space.

Plotting the reachable CM cost contours reveals an extraordinary feature of the 3-dimensional CM cost volume: there are slices through it with CM cost radially symmetric with respect to the origin! The primary effect on the topology of the reachable CM cost surface is that the

four distinct minima merge into a single circular ring on the reachable subspace. Figure 5 demonstrates this claim by use of the rotation angles $\{0, \pi/4, -\cos^{-1}(\sqrt{2/3})\}$.

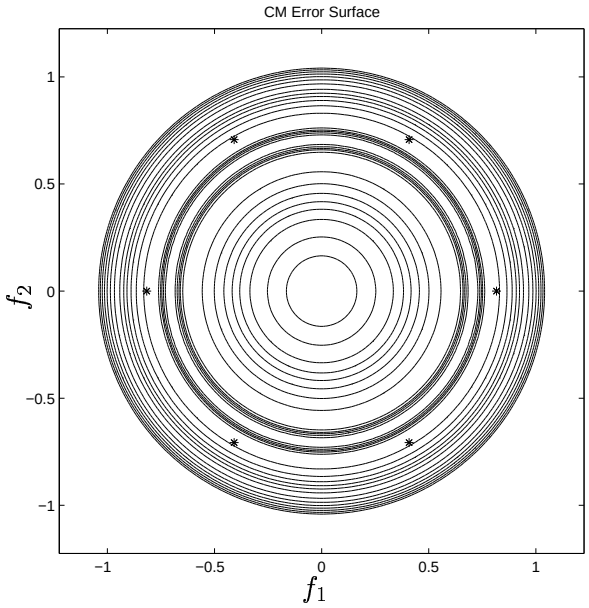

Figure 5: CM cost contours resulting from rotation angles $\{0, \pi/4, -\cos^{-1}(\sqrt{2/3})\}$ in $(\mathbf{p}_1 =)\mathbf{f}$-space.

7 Summary

Two of the original goals of this work have not yet been achieved:

1. counting the number of stationary points (e.g. via zeroing the gradient),
2. counting the number of minima (e.g. via conditions on the Hessian), which should be a channel-dependent quantity.

The resolution of these items are still of great practical interest.

On the other hand, the study so far has uncovered interesting properties about the topology of the CMA cost function in system ($\mathbf{q}$) space. Namely, that there is an unusually high degree of symmetry in certain subspaces of system space (as evidenced by Section 6.2.3). Thus we are encouraged to follow up on the implications and existance of such symmetries in higher dimensions, and to define a class of channels which excite such strange (and unfortunate) behavior.

References

[1] C.R. Johnson, Jr. and B.D.O. Anderson, "Godard blind equalizer error surface charateristics: White, zero-mean, binary case," *IJACSP*, vol. 9, pp. 301-324, 1995.

[2] T.J. Endres, "Equalizing with fractionally-spaced constant modulus and second-order-statistics blind receivers," *Ph.D. Thesis*, Cornell University, Ithaca, NY, 1997.

A Construction of Rotation Matrices

Here we present a technique to construct N-dimensional (orthogonal) rotation matrices parameterized by the angles of rotation between consecutive pairs of coordinate axes. Specifically, angle ϕ_n for $1 \geq n \geq N-1$ specifies the rotation that takes place between axes $\mathbf{e}_n$ and $\mathbf{e}_{n+1}$, while ϕ_N rotates $\mathbf{e}_N$ to $\mathbf{e}_1$. As the order of rotations is important, we adopt the ordering: $\phi_1 \rightarrow \phi_N$.

The rotation matrix $\mathbf{U}$ can be constructed as the product of N rotation matrices $\mathbf{P}_n(\phi_n)$:

$$\mathbf{U}(\phi_1, \ldots, \phi_N) \quad = \quad \mathbf{P}_N(\phi_N) \cdots \mathbf{P}_1(\phi_1) \tag{17}$$

where $\mathbf{P}_n(\cdot)$ effects a rotation between only two dimensions. For $1 \geq n \geq N-1$, we have

$$\mathbf{P}_n(\phi_n) = \begin{pmatrix} \mathbf{I}_{n-1} & & & \\ & \cos\phi_n & -\sin\phi_n & \\ & \sin\phi_n & \cos\phi_n & \\ & & & \mathbf{I}_{N-n-1} \end{pmatrix},$$

and for $\mathbf{P}_N(\phi_N)$ we have

$$\mathbf{P}_N(\phi_N) = \begin{pmatrix} \cos\phi_N & & \sin\phi_N \\ & \mathbf{I}_{N-2} & \\ -\sin\phi_N & & \cos\phi_N \end{pmatrix}.$$