CHAPTER THREE

PRELIMINARY RESULTS FOR PARAFAC WITH ORTHOGONALITY CONSTRAINTS

3.1 INTRODUCTION

In this chapter I discuss some preliminary results pertaining to the PARAFAC model with orthogonality constraints, which will be henceforth referred to as PARAFAC (orth.). These results shall be important for both Chapter Four, which is on Common Principal Components, and Chapter Six, which is on using three-mode methods for CCA, CVA, RA and PR over time. In Section **3.2** I show certain properties of the PARAFAC (orth.) model related to the sums of squares of fit and error. These properties will allow me in Chapter Six to show that the CVA, CCA, RA and PR/time models are optimal in maximizing sums of squared correlations, variance explained, or sums of squared covariances, depending on the method. In Section **3.3** I prove that the PARAFAC (orth.) solutions are nested.

3.2 OPTIMALITY PROPERTIES OF THE PARAFAC MODEL WITH ORTHOGONALITY CONSTRAINTS

Kroonenberg (1983) shows that the least squares solutions to the Tucker2 and Tucker3 models have certain useful properties with respect to the sums of squares. He shows that the total sums of squares can be partitioned into sums of squares fit and sums of squares residuals. He also shows that the sums of the squared elements of the core matrices equal the sums of squares fit. In particular, the square of the i^{th}, j^{th}, k^{th} core element is the fit contributed by the combination of the i^{th} element of the first (subject) mode, the j^{th} element of the second (variable) mode and the k^{th} element of the third (occasion) mode. An important consequence is that minimizing the sums of squares lack of fit is equivalent to maximizing the sums of squares fit, and thus equivalent to maximizing the sums of squares of the elements of the core matrix. In this section I show that these properties hold true for the PARAFAC (orth.) model. Of particular importance to future developments is **Proposition 3.2**.

I start with some necessary definitions. Define $\underline{\mathbf{C}}$ to be an $m \times n \times p$ three-way array, $\mathbf{C}_k = \underline{\mathbf{C}}[,,k]$, $k = 1, \ldots, p$, to be the p $m \times n$ slices of $\underline{\mathbf{C}}$, **F** to be an $m \times m$ orthogonal matrix, **G** to be an $n \times n$ orthogonal matrix, and $\mathbf{D}_k$ to be the $m \times n$ matrix $\mathbf{D}_k = \mathbf{F}'\mathbf{C}_k\mathbf{G}$. Further define $\mathbf{D}_k[i, j]$ to be the element in the i^{th} row and j^{th} column of $\mathbf{D}_k$, $\mathbf{f}_i$ to be the i^{th} column of **F**, and $\mathbf{g}_j$ to be the j^{th} column of **G**.

I proceed with two identities. First, one has the following decomposition of $\mathbf{C}_k$:

$$\mathbf{C}_k = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{g}_i \mathbf{f}_j' \mathbf{D}_k[i, j].$$

Second, the total sums of squares of $\mathbf{D}_k$ equals the total sums of squares of $\mathbf{C}_k$:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{C}_k^2[i, j] = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{D}_k^2[i, j], \tag{3.1}$$

since

$$\text{trace}(\mathbf{C}_k'\mathbf{C}_k) = \text{trace}(\mathbf{G}'\mathbf{C}_k'\mathbf{F}\mathbf{F}'\mathbf{C}_k\mathbf{G}) = \text{trace}(\mathbf{D}_k'\mathbf{D}_k).$$

Define a rank-one approximation to $\mathbf{C}_k$ as $\hat{\mathbf{C}}_k = \mathbf{f}_i\mathbf{g}_j' h_{ij}$. Note that by a theorem by Penrose (1955) for given normal $\mathbf{f}_i$ (i.e., $\|\mathbf{f}_i\|^2 = 1$) and $\mathbf{g}_j$, the optimal h_{ij} is $h_{ij} = \mathbf{f}_i'\mathbf{C}_k\mathbf{g}_j = \mathbf{D}_k[i, j]$. Let T be a non-zero set of combinations of i,j, $1 \le i \le m$, and $1 \le j \le n$. Then the sum of the rank-one approximations to $\mathbf{C}_k$ defined by $\mathbf{f}_i$ and $\mathbf{g}_j$, $(i, j) \in T$, and $h_{ij} = \mathbf{D}_k[i, j]$, is itself an approximation to $\mathbf{C}_k$. Define this as $\hat{\mathbf{C}}_k^T = \sum_{(i,j)\in T} \mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i, j]$.

Proposition 3.1. For the modeling of $\mathbf{C}_k$ by $\hat{\mathbf{C}}_k^T$, the sums of squares total is additively partitioned into the sums of squares fit and sums of squares lack of fit.

Proof: The sums of squares fit is

$$\sum_{(i,j)\in T}(\hat{\mathbf{C}}_k^T[i,j])^2 = \text{trace}(\hat{\mathbf{C}}_k^{T\prime}\hat{\mathbf{C}}_k^T) \quad = \text{trace}\left(\sum_{(i,j)\in T}\mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i,j]\right)'\left(\sum_{(i,j)\in T}\mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i,j]\right)$$

$$= \sum_{(i,j)\in T}\mathbf{D}_k^2[i,j]\text{trace}\left(\mathbf{f}_i\mathbf{g}_j'\right)'\mathbf{f}_i\mathbf{g}_j'$$

$$= \sum_{(i,j)\in T}\mathbf{D}_k^2[i,j]. \tag{3.2}$$

The sums of squares lack of fit is:

$$= \text{trace}(\mathbf{C}_k - \sum_{(i,j)\in T}\mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i,j])'(\mathbf{C}_k - \sum_{(i,j)\in T}\mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i,j])$$

$$= \text{trace}(\mathbf{C}_k'\mathbf{C}_k) + \sum_{(i,j)\in T}\text{trace}(\mathbf{f}_i\mathbf{g}_j'\mathbf{g}_j\mathbf{f}_i'\mathbf{D}_k^2[i,j]) - \sum_{(i,j)\in T}2\text{trace}(\mathbf{C}_k'\mathbf{f}_i\mathbf{g}_j'\mathbf{D}_k[i,j])$$

$$= \text{trace}(\mathbf{C}_k'\mathbf{C}_k) + \sum_{(i,j)\in T}\mathbf{D}_k^2[i,j] - \sum_{(i,j)\in T}2\mathbf{D}_k[i,j]\text{trace}(\mathbf{G}\mathbf{D}_k\mathbf{F}'\mathbf{f}_i\mathbf{g}_j')$$

$$= \text{trace}(\mathbf{C}_k'\mathbf{C}_k) + \sum_{(i,j)\in T}\mathbf{D}_k^2[i,j] - \sum_{(i,j)\in T}2\mathbf{D}_k[i,j]\text{trace}(\mathbf{g}_j'\mathbf{G}\mathbf{D}_k'[i,])$$

$$= \text{trace}(\mathbf{C}_k'\mathbf{C}_k) - \sum_{(i,j)\in T}\mathbf{D}_k^2[i,j].$$

Clearly, the sums of squares fit and sums of squares lack of fit equal the sums of squares total. Ž

For a given **F**, **G** and $\mathbf{C}_k$, $k = 1,\dots,p$, one can consider the class of $\hat{\mathbf{C}}_k^T$, $(i,j)\in T$, as a class of models. I shall refer to these as orthogonal models. Thus by **Proposition 3.1** for an orthogonal model the sums of squares total can be partitioned into a sums of squares fit and a sums of squares error.

With **Proposition 3.1** in place I move to **Proposition 3.2**. Denote the rank-r PARAFAC (orth.) solution to $\mathbf{C}_k$ as $\mathbf{F}^*$, $\mathbf{G}^*$ and $\mathbf{H}^*$, where $\mathbf{F}^*$ is an $m\times r$ columnwise orthonormal matrix, $\mathbf{G}^*$ is an $n\times r$ columnwise orthonormal matrix and $\mathbf{H}^*$ a $p\times r$ diagonal matrix, $r\le m,n,p$. $\mathbf{H}^*$ can also be expressed as p $r\times r$ diagonal matrices $\mathbf{H}_k$, where $k = 1,\dots,p$. This is the form of $\mathbf{H}^*$ which will be used below. Recall it is known (Kroonenberg 1983) that $\mathbf{H}_k = \text{diag}(\mathbf{F}^{*\prime}\mathbf{C}_k\mathbf{G}^*)$.

Proposition 3.2. The least squares estimates of PARAFAC (orth.) are $\mathbf{F}^*$ and $\mathbf{G}^*$ such that $\sum_{k=1}^{g} \text{trace}\left(\mathbf{H}_k^2\right)$ is maximized.

Proof: By definition PARAFAC (orth.) minimizes the total sums of squares lack of fit. Since the PARAFAC (orth.) is in the class of orthogonal models as defined above, by **Proposition 3.1** this is equivalent to maximizing the sums of squares fit. Further, by (3.2) $\sum_{k=1}^{g} \text{trace}\left(\mathbf{H}_k^2\right)$ represents that sums of squares fit. Ž

3.3 THE NESTEDNESS PROPERTY OF PARAFAC SOLUTIONS WITH ORTHOGONALITY CONSTRAINTS

A solution is called nested if the rank-$(f-1)$ solution is a subset of the rank-*f* solution, for any realizable *f*. This implies that one can find any rank-*f* solution recursively by finding *f* rank-one solutions. An example of the nestedness of solutions is the singular value decomposition (Eckart & Young) of a real matrix. A property of real matrices is that the least squares rank-*p* approximation to a matrix can be found by determining *p* rank-one approximations. For example, the best rank-two approximation of a matrix is the sum of its rank-one approximation plus the rank-one approximation to the matrix obtained by subtracting the first rank-one approximation from the original matrix.

The nestedness property is important to modeling with PARAFAC (orth.) because it allows for straightforward comparisons between solutions of different ranks, enabling one to examine the fit attributable to each component. For example, without the nestedness property the component of a rank-one solution may bear no relation to the components of a rank-two solution.

A limited result pertaining to the nestedness of PARAFAC solutions has already been achieved by Leurgans and Ross (1992). They show that a necessary condition for the existence of the nestedness property for a PARAFAC model with an exact rank-two solution is that at least two of the three pairs of components be orthonormal. The following result goes further in that it states that a sufficient condition for the existence of the nestedness property for a PARAFAC model of any rank and of imperfect fit is that two of the three matrices of components be orthonormal.

The subsequent proof will take advantage of conditional linearity. The property of conditional linearity is said to exist if the set of all parameters can be divided into subsets such that in the model each subset is linear in terms of the rest. In the case of the PARAFAC (orth.) model one such division would be the matrices of parameters **F**, **G** and **H**. Because any given subset of parameters is part of the optimal least squares solution, the estimates for those parameters must be the solution to the regression problem where the estimates for the rest of the parameters are treated as fixed. For example, the estimate for **G** must be the solution to the regression problem where $\underline{\mathbf{X}}$ is the response, and **H** and **C** are fixed in the model. This fact is called conditional linearity. Conditional linearity is the basis for the alternating least squares algorithms that are used to obtain least squares estimates for the three-mode models.

As a preliminary I will define trilinear notation that will simplify the presentation. Let **a**, **b** and **c** be vectors. Then the trilinear multiplication of **a**, **b** and **c**, denoted as $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$, is the outer product of **a**, **b** and **c** and generates a three-way array, $\underline{\mathbf{U}}$, such that $\underline{\mathbf{U}}[\mathrm{i}, \mathrm{j}, \mathrm{k}] = \mathrm{a}_{\mathrm{i}} \mathrm{b}_{\mathrm{j}} \mathrm{c}_{\mathrm{k}}$, where a_{i}, b_{j}, and c_{k} are the i^{th}, j^{th} and k^{th} elements of **a**, **b** and **c**. Another way to think of $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$ is that the k^{th} slice of $\underline{\mathbf{U}}$ is $\mathbf{ab}'$ multiplied by c_{k}: $\underline{\mathbf{U}}[,,\mathrm{k}] = \mathbf{ab}'\mathrm{c}_{\mathrm{k}}$.

Theorem 3.1: Let $\underline{\mathbf{X}}$ be a $p \times q \times m$ array. Assume **A**, **B** and **C** represent a rank-*f* PARAFAC solution to $\underline{\mathbf{X}}$ with **A** and **B** constrained to be columnwise orthonormal and the columns of **A**, **B** and **C** ordered by the norm of $\mathbf{c}_{\mathrm{k}}$. That is: $\hat{\underline{\mathbf{X}}} = \mathbf{a}_1 \times \mathbf{b}_1 \times \mathbf{c}_1 + \mathbf{a}_2 \times \mathbf{b}_2 \times \mathbf{c}_2 + \ldots + \mathbf{a}_f \times \mathbf{b}_f \times \mathbf{c}_f$, where $\mathbf{a}_{\mathrm{g}} = \mathbf{A}[,\mathrm{g}]$, $\mathbf{b}_{\mathrm{g}} = \mathbf{B}[,\mathrm{g}]$, and $\mathbf{c}_{\mathrm{g}} = \mathbf{C}[,\mathrm{g}]$ for $\mathrm{g} = 1, \ldots, f$, and $\|\mathbf{c}_{\mathrm{k}}\|^2 \le \|\mathbf{c}_{\mathrm{k}'}\|^2$ for $1 \le \mathrm{k} < \mathrm{k}' \le f$. Then the best rank-*d* solution, $1 \le d \le f$, is $\mathbf{a}_1 \times \mathbf{b}_1 \times \mathbf{c}_1 + \mathbf{a}_2 \times \mathbf{b}_2 \times \mathbf{c}_2 + \ldots + \mathbf{a}_d \times \mathbf{b}_d \times \mathbf{c}_d$.

Proof: Let $\hat{\mathbf{a}}_1 \times \hat{\mathbf{b}}_1 \times \hat{\mathbf{c}}_1 + \hat{\mathbf{a}}_2 \times \hat{\mathbf{b}}_2 \times \hat{\mathbf{c}}_2 + \ldots + \hat{\mathbf{a}}_d \times \hat{\mathbf{b}}_d \times \hat{\mathbf{c}}_d$ be the rank-*d* estimates. Start with the conditional linearity of these estimates. The estimate for any $\hat{\mathbf{a}}_{\mathrm{e}}$, $1 \le \mathrm{e} \le d$, must be the solution to the regression problem where $\underline{\mathbf{X}}$ is the response and $\hat{\mathbf{a}}_{\mathrm{h}}$, $\mathrm{h} = 1, \ldots, d$, $\mathrm{h} \ne \mathrm{e}$, and $\hat{\mathbf{b}}_{\mathrm{h}}$ and $\hat{\mathbf{c}}_{\mathrm{h}}$, $\mathrm{h} = 1, \ldots, d$, are fixed. This is seen below in (3.3), where $\hat{\mathrm{a}}_{\mathrm{ei}}$ denotes the i^{th} element of the vector $\hat{\mathbf{a}}_{\mathrm{e}}$. This regression yields the normal equations given in (3.4) or (3.5). Likewise one can consider $\hat{\mathbf{a}}_{\mathrm{e}}$ and $\hat{\mathbf{b}}_{\mathrm{e}}$ to be fixed and solve for $\hat{\mathbf{c}}_{\mathrm{e}}$ (3.6), and one can consider $\hat{\mathbf{a}}_{\mathrm{e}}$ and $\hat{\mathbf{c}}_{\mathrm{e}}$ to be fixed and solve for $\hat{\mathbf{b}}_{\mathrm{e}}$ (3.7). The least squares solution must simultaneously solve these three regressions.

$$\hat{\mathrm{a}}_{ei} \begin{pmatrix} \hat{\mathrm{b}}_{e1}\hat{\mathrm{c}}_{e1} \\ \vdots \\ \hat{\mathrm{b}}_{eq}\hat{\mathrm{c}}_{e1} \\ \vdots \\ \hat{\mathrm{b}}_{e1}\hat{\mathrm{c}}_{em} \\ \vdots \\ \hat{\mathrm{b}}_{eq}\hat{\mathrm{c}}_{em} \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{X}}[\mathrm{i},1,1] \\ \vdots \\ \underline{\mathbf{X}}[\mathrm{i},q,1] \\ \vdots \\ \underline{\mathbf{X}}[\mathrm{i},1,m] \\ \vdots \\ \underline{\mathbf{X}}[\mathrm{i},q,m] \end{pmatrix} - \sum_{\substack{s=1 \\ s \ne e}}^{d} \hat{\mathrm{a}}_{q\mathrm{i}} \begin{pmatrix} \hat{\mathrm{b}}_{s1}\hat{\mathrm{c}}_{s1} \\ \vdots \\ \hat{\mathrm{b}}_{sq}\hat{\mathrm{c}}_{s1} \\ \vdots \\ \hat{\mathrm{b}}_{s1}\hat{\mathrm{c}}_{sm} \\ \vdots \\ \hat{\mathrm{b}}_{sq}\hat{\mathrm{c}}_{sm} \end{pmatrix}. \tag{3.3}$$

The normal equations for solving for $\hat{\mathrm{a}}_{\mathrm{ei}}$, for $\mathrm{i} = 1, \ldots, p$, are:

$$\sum_{\mathrm{j,k}} \left(\underline{\mathbf{X}}[\mathrm{i}, \mathrm{j}, \mathrm{k}] - \hat{\mathrm{a}}_{1\mathrm{i}} \hat{\mathrm{b}}_{1\mathrm{j}} \hat{\mathrm{c}}_{1\mathrm{k}} - \ldots - \hat{\mathrm{a}}_{\mathrm{ei}} \hat{\mathrm{b}}_{\mathrm{ej}} \hat{\mathrm{c}}_{\mathrm{ek}} \right) \hat{\mathrm{b}}_{\mathrm{ej}} \hat{\mathrm{c}}_{\mathrm{ek}} = 0 \tag{3.4}$$

or in vector form:

$$\sum_{j,k}\left(\underline{\mathbf{X}}[,j,k]-\hat{\mathbf{a}}_1\hat{b}_{1j}\hat{c}_{1k}-\ldots-\hat{\mathbf{a}}_e\hat{b}_{ej}\hat{c}_{ek}\right)\hat{b}_{ej}\hat{c}_{ek}=\mathbf{0}. \tag{3.5}$$

Likewise, one has normal equations for solving for $\hat{\mathbf{b}}_e$ conditioned on $\hat{\mathbf{a}}_e$ and $\hat{\mathbf{c}}_e$ being fixed:

$$\sum_{i,k}\left(\underline{\mathbf{X}}[i,,k]-\hat{a}_{1i}\hat{\mathbf{b}}_1\hat{c}_{1k}-\ldots-\hat{a}_{ei}\hat{\mathbf{b}}_e\hat{c}_{ek}\right)\hat{a}_{ei}\hat{c}_{ek}=\mathbf{0} \tag{3.6}$$

and one has normal equations for solving for $\hat{\mathbf{c}}_e$ conditioned on $\hat{\mathbf{a}}_e$ and $\hat{\mathbf{b}}_e$ being fixed:

$$\sum_{i,j}\left(\underline{\mathbf{X}}[i,j,]-\hat{a}_{1i}\hat{b}_{1j}\hat{\mathbf{c}}_1-\ldots-\hat{a}_{ei}\hat{b}_{ej}\hat{\mathbf{c}}_e\right)\hat{a}_{ei}\hat{b}_{ej}=\mathbf{0}. \tag{3.7}$$

Let $\underline{\mathbf{E}}=\underline{\mathbf{X}}-\hat{\underline{\mathbf{X}}}$. Then the normal equations for $\hat{\mathbf{a}}_e$, $\hat{\mathbf{b}}_e$ and $\hat{\mathbf{c}}_e$ can be written as

$$\sum_{j,k}(\mathbf{a}_1 b_{1j}c_{1k}+\mathbf{a}_2 b_{2j}c_{2k}+\cdots+\mathbf{a}_f b_{fj}c_{fk}+\underline{\mathbf{E}}[,j,k]-\hat{\mathbf{a}}_1\hat{b}_{1j}\hat{c}_{1k}-\ldots-\hat{\mathbf{a}}_e\hat{b}_{ej}\hat{c}_{ek})\hat{b}_{ej}\hat{c}_{ek}=\mathbf{0},$$

$$\sum_{i,k}(a_{1i}\mathbf{b}_1 c_{1k}+a_{2i}\mathbf{b}_2 c_{2k}+\cdots+a_{fi}\mathbf{b}_f c_{fk}+\underline{\mathbf{E}}[i,,k]-\hat{a}_{1i}\hat{\mathbf{b}}_1\hat{c}_{1k}-\ldots-\hat{a}_{ei}\hat{\mathbf{b}}_e\hat{c}_{ek})\hat{a}_{ei}\hat{c}_{ek}=\mathbf{0},$$

$$\sum_{i,j}(a_{1i}b_{1j}\mathbf{c}_1+a_{2i}b_{2j}\mathbf{c}_2+\cdots+a_{fi}b_{fj}\mathbf{c}_f+\underline{\mathbf{E}}[i,j,]-\hat{a}_{1i}\hat{b}_{1j}\hat{\mathbf{c}}_1-\ldots-\hat{a}_{ei}\hat{b}_{ej}\hat{\mathbf{c}}_e)\hat{a}_{ei}\hat{b}_{ej}=\mathbf{0}.$$

Collecting terms in j and k allows these normal equations to be written as follows:

$$\mathbf{a}_1\left(\sum_j b_{1j}\hat{b}_{ej}\right)\left(\sum_k c_{1k}\hat{c}_{ek}\right)+\cdots+\mathbf{a}_f\left(\sum_j b_{fj}\hat{b}_{ej}\right)\left(\sum_k c_{fk}\hat{c}_{ek}\right)$$
$$-\hat{\mathbf{a}}_1\left(\sum_j \hat{b}_{1j}\hat{b}_{ej}\right)\left(\sum_k \hat{c}_{1k}\hat{c}_{ek}\right)-\ldots-\hat{\mathbf{a}}_e\left(\sum_j \hat{b}_{ej}\hat{b}_{ej}\right)\left(\sum_k \hat{c}_{ek}\hat{c}_{ek}\right)+\sum_{j,k}\underline{\mathbf{E}}[,j,k]\hat{b}_{ej}\hat{c}_{ek}=\mathbf{0} \tag{3.8}$$

$$\mathbf{b}_1\left(\sum_i a_{1j}\hat{a}_{ei}\right)\left(\sum_k c_{1k}\hat{c}_{ek}\right)+\cdots+\mathbf{b}_f\left(\sum_i a_{fi}\hat{a}_{ei}\right)\left(\sum_k c_{fk}\hat{c}_{ek}\right)$$
$$-\hat{\mathbf{b}}_1\left(\sum_i \hat{a}_{1i}\hat{a}_{ei}\right)\left(\sum_k \hat{c}_{1k}\hat{c}_{ek}\right)-\ldots-\hat{\mathbf{b}}_e\left(\sum_i \hat{a}_{ei}\hat{a}_{ei}\right)\left(\sum_k \hat{c}_{ek}\hat{c}_{ek}\right)+\sum_{i,k}\underline{\mathbf{E}}[i,,k]\hat{a}_{ei}\hat{c}_{ek}=\mathbf{0} \tag{3.9}$$

$$\mathbf{c}_1\left(\sum_i a_{1i}\hat{a}_{ei}\right)\left(\sum_j b_{1j}\hat{b}_{ej}\right)+\cdots+\mathbf{c}_f\left(\sum_i a_{fi}\hat{a}_{ei}\right)\left(\sum_j b_{fj}\hat{b}_{ej}\right)$$
$$-\hat{\mathbf{c}}_1\left(\sum_i \hat{a}_{1i}\hat{a}_{ei}\right)\left(\sum_j \hat{b}_{1j}\hat{b}_{ej}\right)-\ldots-\hat{\mathbf{c}}_e\left(\sum_i \hat{a}_{ei}\hat{a}_{ei}\right)\left(\sum_j \hat{b}_{ej}\hat{b}_{ej}\right)+\sum_{i,j}\underline{\mathbf{E}}[i,j,]\hat{a}_{ei}\hat{b}_{ej}=\mathbf{0}. \tag{3.10}$$

Several of the terms in equations 3.8, 3.9 and 3.10 drop out. By the definition of the normal equations $\sum_{j,k}\underline{\mathbf{E}}[,j,k]\hat{b}_{ej}\hat{c}_{ek}=\sum_{i,k}\underline{\mathbf{E}}[i,,k]\hat{a}_{ej}\hat{c}_{ek}=\sum_{i,j}\underline{\mathbf{E}}[i,j,]\hat{a}_{ei}\hat{b}_{ek}=\mathbf{0}$. Because of the orthonormality of $\hat{\mathbf{a}}_g$ and of $\hat{\mathbf{b}}_g$, $g=1,\ldots,d$, one has $\left(\sum_i \hat{a}_{gi}\hat{a}_{ei}\right)=\left(\sum_j \hat{b}_{gj}\hat{b}_{ej}\right)=0$ if $g\neq e$ or 1 if $g=e$.

Consider that $\sum_i a_{mi}\hat{a}_{ni} = \text{cosine}(\alpha_{mn})$, where α_{mn} is the angle subtended between $\mathbf{a}_m$ and $\hat{\mathbf{a}}_n$, and $\sum_j b_{mj}\hat{b}_{nj} = \text{cosine}(\beta_{mn})$, where β_{mn} is the angle subtended between $\mathbf{b}_m$ and $\hat{\mathbf{b}}_n$. Then from (3.10) $\hat{\mathbf{c}}_e$ is seen to be a weighted sum of $\mathbf{c}_1,\ldots,\mathbf{c}_f$:

$$\hat{\mathbf{c}}_e = \mathbf{c}_1 \text{cosine}(\alpha_{1e})\text{cosine}(\beta_{1e}) + \mathbf{c}_2 \text{cosine}(\alpha_{2e})\text{cosine}(\beta_{2e}) + \ldots + \mathbf{c}_f \text{cosine}(\alpha_{fe})\text{cosine}(\beta_{fe}). \quad (3.11)$$

Now, the relation in (3.11) is true for $e = 1,\ldots,d$, thus one has

$$\hat{\mathbf{c}}_1 = \mathbf{c}_1 \text{cosine}(\alpha_{11})\text{cosine}(\beta_{11}) + \mathbf{c}_2 \text{cosine}(\alpha_{21})\text{cosine}(\beta_{21}) + \ldots + \mathbf{c}_f \text{cosine}(\alpha_{f1})\text{cosine}(\beta_{f1})$$
$$\hat{\mathbf{c}}_2 = \mathbf{c}_1 \text{cosine}(\alpha_{12})\text{cosine}(\beta_{12}) + \mathbf{c}_2 \text{cosine}(\alpha_{22})\text{cosine}(\beta_{22}) + \ldots + \mathbf{c}_f \text{cosine}(\alpha_{f2})\text{cosine}(\beta_{f2})$$
$$\vdots$$
$$\hat{\mathbf{c}}_d = \mathbf{c}_1 \text{cosine}(\alpha_{1d})\text{cosine}(\beta_{1d}) + \mathbf{c}_2 \text{cosine}(\alpha_{2d})\text{cosine}(\beta_{2d}) + \ldots + \mathbf{c}_f \text{cosine}(\alpha_{fd})\text{cosine}(\beta_{fd}).$$

Clearly the upper bound for $\sum_{k=1}^{d}\|\hat{\mathbf{c}}_k\|^2$ is $\sum_{k=1}^{d}\|\mathbf{c}_k\|^2$, and this is achieved when $\text{cosine}(\alpha_{11}) = \text{cosine}(\beta_{11}) = 1$, $\text{cosine}(\alpha_{22}) = \text{cosine}(\beta_{22}) = 1$, …, and $\text{cosine}(\alpha_{ee}) = \text{cosine}(\beta_{ee}) = 1$, or when $\hat{\mathbf{a}}_1 = \mathbf{a}_1$, $\hat{\mathbf{b}}_1 = \mathbf{b}_1$, $\hat{\mathbf{a}}_2 = \mathbf{a}_2$, $\hat{\mathbf{b}}_2 = \mathbf{b}_2$,..., $\hat{\mathbf{a}}_d = \mathbf{a}_d$, and $\hat{\mathbf{b}}_d = \mathbf{b}_d$. Note this solution satisfies the normal equations, as $\sum_i a_{gi}a_{ei} = 0$ and $\sum_j b_{gj}b_{ej} = 0$ for $e \neq g$. Now by **Proposition 3.2**, the PARAFAC (orth.) solution is **A** and **B** such that the sums of squares of $\sum_{k=1}^{d}\|\hat{\mathbf{c}}_k\|^2$ is maximized. Since the solution $\hat{\mathbf{a}}_1 = \mathbf{a}_1$, $\hat{\mathbf{b}}_1 = \mathbf{b}_1$, $\hat{\mathbf{a}}_2 = \mathbf{a}_2$, $\hat{\mathbf{b}}_2 = \mathbf{b}_2$,..., $\hat{\mathbf{a}}_d = \mathbf{a}_d$, $\hat{\mathbf{b}}_d = \mathbf{b}_d$, and $\hat{\mathbf{c}}_1 = \mathbf{c}_1$, $\hat{\mathbf{c}}_2 = \mathbf{c}_2$,..., $\hat{\mathbf{c}}_d = \mathbf{c}_d$, maximizes this sums of squares and satisfies the least squares normal equations, it is the rank-*d* solution.