A NEW EXPLANATION AND PROOF OF THE PARADOXICAL SCORING RESULTS IN MULTIDIMENSIONAL ITEM RESPONSE MODELS

PASCAL JORDAN AND MARTIN SPIESS

UNIVERSITY OF HAMBURG

In multidimensional item response models, paradoxical scoring effects can arise, wherein correct answers are penalized and incorrect answers are rewarded. For the most prominent class of IRT models, the class of linearly compensatory models, a general derivation of paradoxical scoring effects based on the geometry of item discrimination vectors is given, which furthermore corrects an error in an established theorem on paradoxical results. This approach highlights the very counterintuitive way in which item discrimination parameters (and also factor loadings) have to be interpreted in terms of their influence on the latent ability estimate. It is proven that, despite the error in the original proof, the key result concerning the existence of paradoxical effects remains true—although the actual relation to the item parameters is shown to be a more complicated function than previous results suggested. The new proof enables further insights into the actual mathematical causation of the paradox and generalizes the findings within the class of linearly compensatory models.

Key words: paradoxical result, test fairness, compensatory model, multidimensional item response theory, person parameter estimation, hyperplanes.

1. Introduction

In 2009 Hooker, Finkelman and Schwartzman established a paradoxical effect within the class of binary multidimensional item response models. By a paradoxical effect the authors refer to the case that changing a correct response on an item into an incorrect response results in an increase in the estimate (typically the maximum likelihood estimate) of at least one latent ability. This effect does not occur when simple structure[1] is present, but is an inherent part of "truly" multidimensional scales. The authors (Hooker, Finkelman and Schwartzman) not only highlighted the potentially severe practical consequences of this effect with respect to classification decisions and test fairness, but also provided an in depth analysis of the effect encompassing a variety of estimation procedures (e.g., including Bayesian estimates). They developed a sophisticated proof which demonstrated a constructive method to generate a paradoxical result under a rather general model framework. The authors also identified the relevant modeling assumptions which cause the paradox and therefore provided also the basis for further analysis and generalizations of the effect. Subsequent derivations and generalizations of the effect were given by van der Linden (2012), who used properties of level sets of the likelihood function, and by Jordan & Spiess (2012), who used the concept of a reverse rule function (Karlin, 1980). Finally, the phenomenon has been placed into the broader context of Bayesian networks and graphical models as shown by van Rijn & Rijmen (2012). All of these treatments provided a general analysis of paradoxical scoring effects. They did not tackle the omnipresence of the effect in linearly compensatory models—as this issue seemed to have been already addressed by a corresponding theorem in the first derivation by Hooker et al. (2009).

Correspondence should be made to Pascal Jordan, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany. Email: pascal.jordan@uni-hamburg.de

[1]By simple structure we mean that each row of the matrix has exactly one nonzero entry.

The present paper deals with this small—but yet important—part of the paper by Hooker et al. (2009), namely the analysis of the effect in the class of linearly compensatory models. With respect to this class, Hooker et al. (2009) stated that the effect is certain to occur and supplied a proof (Corollary 6.1 of Hooker et al., 2009) which enabled the precise identification of an item inducing paradoxical scoring. The identification is as follows: rank the items with respect to their relative loading on the first latent dimension and then choose the item with the least relative loading on the first dimension. The claim is that changing the item score on this item (e.g., correct into incorrect response or vice versa) always induces a paradoxical effect with respect to the first dimension. Although the result can be substantiated with intuitive reasoning, we will show in this paper that it is not correct in general. More specifically, it is only correct within a two-dimensional test and we will provide a counterexample showing that the above rule to identify an item causing paradoxical movements on the first dimension is false in a three-dimensional setting. Note that this error refers to an important theorem, but that it does not have any consequences with respect to all other results which were derived in Hooker et al. (2009).

Of course, as the proof of the result is false for $p > 2$ dimensions, the claim that the paradox is certain to occur (within higher-dimensional tests) needs to be reevaluated. We will supply a proof showing that the paradox still holds, although the rule of identifying items needs to be qualified. The new proof not only establishes the paradox for higher-dimensional settings but, more importantly, it also generalizes the effect and sheds new light on the mathematical causation of it by furnishing a geometrical interpretation. As a byproduct of the analysis of the paradoxical effect, we will also show, via an example, the incorrectness of the familiar rule expressed e.g., on p. 26 of van der Linden (2012) as:

> "The underlying principle is intuitively clear, though. If an extra item discriminates highly along one of the dimensions relative to the others, an incorrect or correct response to it has a strong downward or upward impact on the MLE for the dimension, respectively. In order to compensate, the MLEs for some of the other dimensions will move in the opposite direction."

Consequently, researchers should be aware of the complicated dependency of the latent ability estimates on the structure of item discrimination.

2. The Modeling Framework

Before tackling the paradoxical scoring effect in depth, we describe the modeling framework to which all subsequent derivations will refer to.

The goal is to infer the unknown values of some latent abilities—abbreviated as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$, wherein the subscript p indicates the dimensionality of the corresponding IRT model—from the responses of the test taker to a set of items with known item parameters.[2] The response of the test taker to the ith item is denoted as U_i. The unknown parameter $\boldsymbol{\theta}$ is related to the responses on the items via the function $P(U_1 = u_1, \ldots U_k = u_k | \boldsymbol{\theta})$, which can be decomposed according to the commonly applied local independence assumption as follows:

$$P(U_1 = u_1, \ldots U_k = u_k | \boldsymbol{\theta}) = \prod_{i=1}^{k} P(U_i = u_i | \boldsymbol{\theta}).$$

---

[2]The assumption of known item parameters is commonly applied and justifiable if the scale has been calibrated with a sufficiently large sample from the population of test takers.

The expression $\widetilde{l}_{i,u_i}(\boldsymbol{\theta}) := \log(P(U_i = u_i|\boldsymbol{\theta}))$ will be used to denote the loglikelihood contribution of the ith item response (for ease of notation the subscript u_i in $\widetilde{l}_{i,u_i}$ will sometimes be suppressed).

One major contribution of the work by Hooker et al. (2009) was the identification of modeling assumptions which provide a fruitful setup for the analysis of the paradox. Apart from some minor changes, we just restate these assumptions as they will, again, play a crucial role in the analysis. That is, the models which will be discussed further on are characterized by the following assumptions (examples are given below):

(a) The loglikelihood contribution $\widetilde{l}_i$ may be written as: $\widetilde{l}_i(\boldsymbol{\theta}) = l_i(\boldsymbol{a}_i^T\boldsymbol{\theta})$ with $\boldsymbol{a}_i^T = (a_{i,1}, \ldots a_{i,p})$ denoting the item discrimination vector of the ith item and l_i denoting a real-valued function with domain $\mathbb{R}$ (for $i = 1, \ldots k$).

(b) Each l_i is twice continuously differentiable with strictly negative second derivative: $l_i'' < 0$.

(c) The $(k \times p)$ matrix of item discrimination vectors A (with ith row equal to $\boldsymbol{a}_i$) is of full column rank p and not of simple structure.

(d) Each item discrimination vector contains only nonnegative entries.

Assumption (a) implies that the equi-probability contours are hyperplanes. That is, for each item, the set of $\boldsymbol{\theta}$ which are indistinguishable with respect to the probability of the response u_i on the item is equal to a hyperplane $\boldsymbol{a}_i^T\boldsymbol{\theta} = c$ with the item discrimination vector $\boldsymbol{a}_i$ acting as the normal vector. Assumption (b) implies [in conjunction with (c)] that the loglikelihood is a strictly concave function which in turn implies that the MLE—if it exists—is unique. Assumption (c) rules out the case that the scale is decomposable into unidimensional scales and states that the scale is p-dimensional. Note that the case of simple structure is the benevolent case, in which no paradoxical scoring effects can occur when using MLEs (although the Bayesian analogue is not necessarily free of paradoxical scoring—see Hooker, 2010). Therefore, we do not treat this special case. Finally, assumption (d) implies that the abilities are compensatory. If, for example, the term l_i refers to a correct response, then l_i is increasing (e.g., $l_i = \log(\frac{\exp(x+\beta_i)}{1+\exp(x+\beta_i)})$) in case of a multidimensional two-parameter logistic model). Hence, the probability of solving an item is a nondecreasing function of each latent ability under assumption (d). Decreases in some latent abilities need not decrease the probability of solving an item if "compensated" by increases on other dimensions.

We shall collectively refer to an IRT model satisfying all of these assumptions (a)–(d) together with the local independence assumption as a linearly compensatory model—compare also with definitions given in Reckase (2009) or van der Linden (2012). In addition, we shall in all subsequent derivations tacitly assume the existence of the MLE, although exact conditions ensuring the existence could be given by using the concept of a recession function (see e.g., ch. 27 of Rockafellar, 1970).

The above description of the modeling class is abstract. Before moving on toward the main topic, we will provide three examples of commonly applied models (see Reckase (2009) for further description) which all fall within the scope of the class of linearly compensatory models (assumptions c) and d) are presupposed in any of the following cases).

- In the *multidimensional two-parameter logistic model* (M2PL) the probability of solving an item is defined as:

$$P(U_i = 1|\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{a}_i^T\boldsymbol{\theta} + \beta_i)}{1 + \exp(\boldsymbol{a}_i^T\boldsymbol{\theta} + \beta_i)}$$

It may be checked by direct computation that the second derivatives of the functions $l_{i,u_i=1}(x) = \log(\frac{\exp(x+\beta_i)}{1+\exp(x+\beta_i)})$ and $l_{i,u_i=0}(x) = \log(1 - \frac{\exp(x+\beta_i)}{1+\exp(x+\beta_i)})$ are negative throughout $\mathbb{R}$.

- In the *multidimensional graded response model* (Samejima, 1974), the probability of obtaining score j on the ith item is defined as:

$$P(U_i = j|\boldsymbol{\theta}) = \Phi(\boldsymbol{a}_i^T\boldsymbol{\theta} + \tau_{i,j}) - \Phi(\boldsymbol{a}_i^T\boldsymbol{\theta} + \tau_{i,j-1})$$

with ordered thresholds $\tau_{i,j} > \tau_{i,j-1}$. The derivation of the condition $l_i'' < 0$ for the function $l_i(x) := \log(\Phi(x + \tau_{i,j}) - \Phi(x + \tau_{i,j-1}))$ is, e.g., given in Jordan & Spiess (2012).

- Up to a proportionality factor, the density of an observed score u_i on the ith item in a *factor analysis model* (FA) is given by:

$$f(u_i|\boldsymbol{\theta}) = \exp\left(-\frac{1}{2\sigma_i^2}(u_i - (\mu_i + \boldsymbol{a}_i^T\boldsymbol{\theta}))^2\right).$$

The term σ_i^2 equals the measurement error variance of the ith item, the term $-\mu_i$ equals the item difficulty and the latent factor $\boldsymbol{f}$ is written as $\boldsymbol{\theta}$. The logarithm of this function equals the ith loglikelihood contribution and may be written as:

$$\widetilde{l}_{i,u_i}(\boldsymbol{\theta}) = \log(f(u_i|\boldsymbol{\theta})) = -\frac{1}{2\sigma_i^2}(u_i - (\mu_i + \boldsymbol{a}_i^T\boldsymbol{\theta}))^2 = l_i(\boldsymbol{a}_i^T\boldsymbol{\theta})$$

with $l_i(x) := -\frac{1}{2\sigma_i^2}(u_i - \mu_i - x)^2$. The first derivative of the latter function is easily computed as: $l_i'(x) = \frac{1}{\sigma_i^2}(u_i - \mu_i - x)$. Therefore, the second derivative equates to $l_i''(x) = -\frac{1}{\sigma_i^2}$ which is negative throughout $\mathbb{R}$.

We emphasize that although the original derivation of Hooker et al. (2009) was concerned with binary MIRT models, the paradox has also been shown to occur in ordinal item response models as well as in factor analysis models (Jordan & Spiess, 2012). Herein, the term "paradoxical effect" refers to the case that increases in item scores decrease some latent abilities. As the reasoning underlying the derivation of our main result also holds for these types of models, we have included them in the above model specification.

Remark. A final note concerns the notion of ordering of ability estimates (here: maximum likelihood estimates—MLEs). Given two different loglikelihood functions f_1, f_2 and their implied MLEs $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, we shall subsequently speak of an ordering of these two MLEs whenever they are ordered in the sense of the usually applied partial ordering in $\mathbb{R}^p$, i.e., $\theta_{1,i} \leq \theta_{2,i}$ for all i and $\theta_{1,j} < \theta_{2,j}$ for at least one dimension j. In case of the analysis of the paradoxical effect, the most frequently applied case is given by specifying the first loglikelihood according to the above model as $f_1 := \sum_{i=1}^{k} \widetilde{l}_{i,u_i}$ and setting the second loglikelihood as equal to the first loglikelihood except for the answer on one specific item (e.g., $f_2 = f_1 + (\widetilde{l}_{k,u_k^*} - \widetilde{l}_{k,u_k})$). With respect to this chosen item (without loss of generality: the last item k), the score is increased and the ordering of the corresponding MLEs is examined. That is, it is examined, whether an increase in an item score leads to increases in all components of the latent ability estimates. If, on the contrary, the two MLEs are not ordered, then we speak of a paradoxical scoring effect. Although this case provides the usual setup for the analysis of the paradox, in our main derivations (see Sect. 3.4) we shall use

a different approach, wherein the first loglikelihood equals $f_1 = \sum_{i=1}^{k-1} \widetilde{l}_{i,u_i}$ and the second equals $f_2 = \sum_{i=1}^{k} \widetilde{l}_{i,u_i}$. Herein, we analyze the ordering of an MLE corresponding to a so-called reduced loglikelihood, wherein the performance on item k does not appear, in comparison with the MLE corresponding to a full loglikelihood, i.e., a loglikelihood also including the performance on the kth item.

3. Ordering of MLEs and Paradoxical Effects Within Linearly Compensatory Models

Subsequently, the notation A_{-i} will be used to indicate the matrix which consists of all item discrimination vectors except the ith and e_1 will denote the first unit vector. It will also be assumed that the matrix A_{-i} still has full column rank p—regardless of the particular item discrimination parameter which is removed, i.e., for any i: $r(A_{-i}) = p$.

One key result in the general derivation of paradoxical results concerns conditions on the item parameters which lead to a paradoxical scoring effect (Theorem 6.1 of Hooker et. al., 2009). Given a linearly compensatory model, a sufficient condition for the existence of a paradoxical result may be phrased as:

Theorem 1. (see Theorem 6.1. of Hooker et al.) *If*

$$e_1^T (A_{-i}^T W A_{-i})^{-1} a_i < 0$$

holds for all positive $(k-1) \times (k-1)$ diagonal matrices W, then changing an incorrect response on the ith item into a correct response will result in a lower estimate for the first latent dimension.

Proof. See Hooker et al. (2009). In Appendix A, we provide an alternative proof using the implicit function theorem (see e.g., Dontchev & Rockafellar, 2009) and a perturbation parameter.

Remark. A similar result holds with respect to the jth latent dimension if one substitutes e_j ($j = 2, \ldots, p$) for e_1 in the above inequality. Without loss of generality the subsequent discussion will only refer to the first latent dimension.

This very useful theorem (see Appendix A for an alternative proof) was used to derive the following rule (Corollary 6.1 in Hooker et al., 2009) which identifies an item causing paradoxical movements with respect to a particular dimension and which at the same time established the presence of paradoxical effects within any linearly compensatory model:

Theorem 2. *(see Corollary 6.1 of Hooker et al.) Choose an item i such that*

$$\frac{a_{i,1}}{|a_i|} < \frac{a_{j,1}}{|a_j|}$$

holds for all $j \neq i$. Then changing a correct response on this item into an incorrect response will increase the estimate for the proficiency on the first dimension.

Remark. Note that the quantity $\frac{a_{j,1}}{|a_j|}$ equals $\frac{e_1^T a_j}{|e_1||a_j|}$, so that the above condition [recall assumption (d)] is equivalent to stating that the discrimination vector a_i represents the vector exhibiting the largest angle with the first coordinate axis.

The first key observation we want to convey is that the rule expressed in Theorem 2 holds only for the special case of a two-dimensional scale. The theorem is, however, false in higher-dimensional settings (thus, also the presence of the paradoxical effect cannot be derived anymore from the theorem). In the following, we will

(1) provide a counterexample showing that the conjecture is false in higher-dimensional settings;

(2) highlight the error in the proof of the theorem;

(3) establish a shortened proof for the two-dimensional setting;

and finally (the main contribution)

(4) provide a modified rule which also works for $p > 2$ and which further allows for a generalization of and new insights into the paradoxical effect.

3.1. A counterexample

We first note that there is nothing inherent in the proof of Hooker et. al. (2009) which necessitates a binary response variable. In fact, all derivations can be adopted for the factor analysis model (see e.g., Jordan & Spiess, 2012) by noting that the loglikelihood corresponding to a response pattern $u_1, u_2, \ldots u_k$ of a factor analysis model (assuming known item parameters) can be written up to a proportionality constant as

$$\sum_i l_{i,u_i}(\boldsymbol{a}_i^T \boldsymbol{\theta})$$

with $l_{i,u_i}(x) := -\frac{1}{2\sigma_i^2}(u_i - \mu_i - x)^2$ satisfying $l_i'' = -\frac{1}{\sigma_i^2} < 0$—just as outlined toward the end of Sect. 2. Thus, assumptions (a) and (b) also hold for the loglikelihood of a factor analysis model (FA).

The counterexample we provide is within the framework of the FA model, because the latter allows for a direct computation of the MLE. The counterexample could also be casted within a binary MIRT model, but we think that a counterexample enabling direct computation is more fruitful than a counterexample relying on computational optimization (for the sake of completeness, a counterexample within the binary MIRT setting is nevertheless given in Appendix B).

Suppose therefore that we are given a FA model[3] with the following matrix of factor loadings (and equal measurement error variances $\sigma_i^2 = 1$):

$$A := \begin{pmatrix} 1 & 1 & 3 \\ 5 & 3 & 1 \\ 2 & 2 & 4 \\ 1 & 1 & 1 \end{pmatrix}.$$

The corresponding matrix of "standardized" (according to the quantity $d_i^j := \frac{a_{ij}}{|\boldsymbol{a}_i|}$ appearing in the rule of Hooker et al., 2009) loadings is then given by

$$D := \begin{pmatrix} d_1^1 = \frac{1}{\sqrt{11}} & d_1^2 = \frac{1}{\sqrt{11}} & d_1^3 = \frac{3}{\sqrt{11}} \\ d_2^1 = \frac{5}{\sqrt{35}} & d_2^2 = \frac{3}{\sqrt{35}} & d_2^3 = \frac{1}{\sqrt{35}} \\ d_3^1 = \frac{1}{\sqrt{6}} & d_3^2 = \frac{1}{\sqrt{6}} & d_3^3 = \frac{2}{\sqrt{6}} \\ d_4^1 = \frac{1}{\sqrt{3}} & d_4^2 = \frac{1}{\sqrt{3}} & d_4^3 = \frac{1}{\sqrt{3}} \end{pmatrix}.$$

---

[3]Note that we are dealing with factor score estimation assuming known factor loadings. As a consequence of this, the unknown person parameter/factor score is identifiable (this requires only a full column rank matrix—as in ordinary linear regression).

As the first item represents the item with the least relative emphasis on the first dimension ($\frac{1}{\sqrt{11}} <$ $\frac{1}{\sqrt{6}} < \frac{1}{\sqrt{3}} < \frac{5}{\sqrt{35}}$), one would expect—according to Theorem 2—that increasing the item score on the first item induces a paradoxical effect on the estimate of the first latent dimension. However, a direct computation of the MLE[4] (p. 274 in Mardia, Kent & Bibby, 1979) gives

$$C := (A^T A)^{-1} A^T = \begin{pmatrix} 1.00 & 0.50 & -0.50 & -1.50 \\ -1.83 & -0.50 & 0.83 & 2.68 \\ 0.50 & 0.00 & 0.00 & -0.50 \end{pmatrix}.$$

Since an entry c_{ij} of the matrix C corresponds to a change of the estimate refering to the ith latent dimension induced by a one unit increase in the jth item score, a contradiction arises: The first item—though representing the item with the least relative emphasis on the first dimension—does not cause a paradoxical movement on the first dimension. In fact, the corresponding change of 1.00 is positive (and doubles the increase corresponding to the third dimension—although the loading relative to the latter dimension equals thrice the loading on the first dimension).

This provides a counterexample to the conjecture of Corollary 6.1. of Hooker et al. (2009) and therefore raises the question as to whether the paradox is also certain to occur in $p > 2$ dimensional settings. Before turning to this crucial point, we will strenghten the implication of the counterexample by highlighting an error in the proof of Corollary 6.1.

3.2. The error in the proof

We can highlight the difference between the two-dimensional and the higher-dimensional setting by geometry. Indeed, the crucial point where the proof of Hooker et al. is not applicable in higher-dimensional settings ($p > 2$) is given by the following statement (see the proof of Lemma D.2 of Hooker et al., 2009):

- If b denotes the discrimination vector with the least relative emphasis on the first dimension (i.e., largest angle to the respective coordinate axis), then the first component of the projection of b onto the orthogonal complement of the subspace spanned by any choice of $p - 1$ linearly independent item discrimination vectors (chosen from the same scale) is nonpositive.

This statement is true in a two-dimensional setting. Although a proof of this is fairly straightforward, we decide to depict this statement via a graphical example—see Fig. 1. However, the choice of

$$b = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \quad a_{s(1)} = \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}, \quad a_{s(2)} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

shows that the statement is false in a three-dimensional setting: the projection of b onto the orthogonal complement of the subspace spanned by[5] $a_{s(1)}$ and $a_{s(2)}$ is given by the expression

$$b - A(A^T A)^{-1} A^T b,$$

[4]For simplicity, we assume equal measurement variance across the items, so that the MLE under normality is identical to the least-squares-estimator.

[5]Note that $s(\cdot)$ describes a mapping from the set $\{1, 2, \ldots p - 1\}$ into the set $\{1, 2, \ldots k\}$—the mapping which selects $\{1, 2, \ldots p - 1\}$ linearly independent vectors from the set of item discrimination vectors.
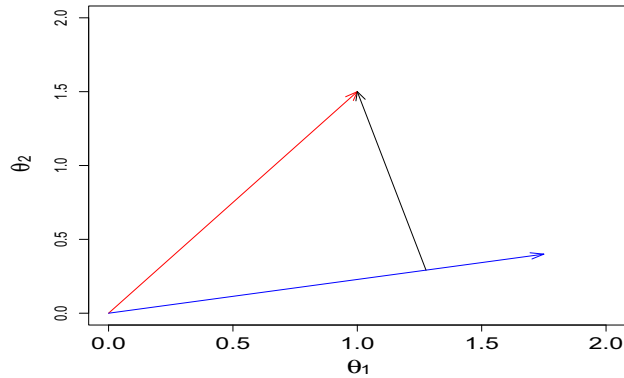
FIGURE 1.
Depicted is the projection of the red vector onto the orthogonal complement of the subspace spanned by the blue vector. As can be seen, the first component of this projection (black vector) is negative. Note that this will always happen if the red vector is further away from the θ_1-axis (or equivalently: more closely aligned to the θ_2-axis) (Color figure online).
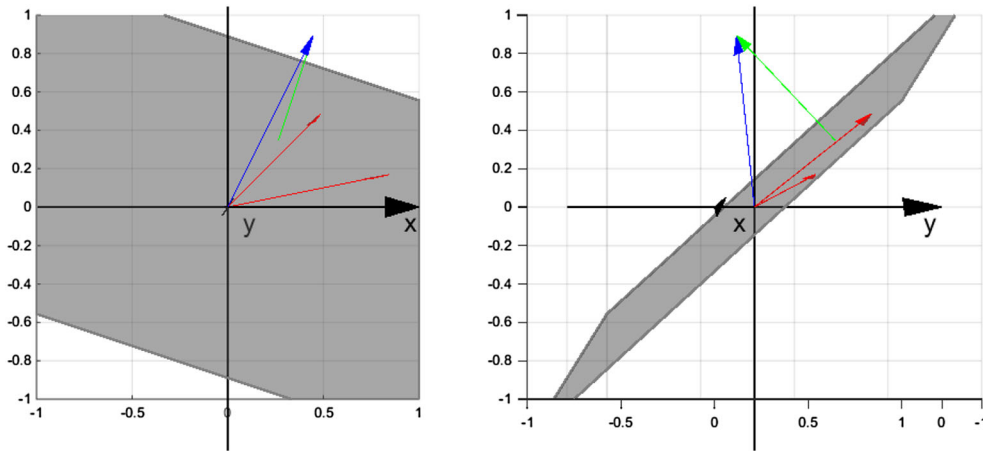

FIGURE 2.
A hyperplane (gray) spanned by two linearly independent vectors (in red). The blue vector lies outside the hyperplane. The corresponding projection onto the orthogonal complement, depicted in green, has a positive x-coordinate—despite the fact that the blue vector exhibits the largest angle with the x-axis (Color figure online).

wherein A represents the 3×2 matrix whose columns are $\boldsymbol{a}_{s(1)}$ and $\boldsymbol{a}_{s(2)}$. A direct computation shows that the first component of this expression is positive (0.09)—despite the fact that $\boldsymbol{b}$ puts the least relative weight on the first dimension.

A graphical illustration of a similar counterexample is further provided in Fig. 2.

Remark. The modified statement that there is some nonpositive component is true, but the exact specification (first component) is false.

3.3. Proof for the Two-Dimensional Setting

This subsection provides a shortened proof of the paradoxical scoring effect in the two-dimensional setting. The reader who is mostly interested in the general mathematical causation

of the paradox and the specific treatment for $p > 2$ dimensional tests can skip it without harm and directly continue with the main part of the paper—located in Sect. 3.4.

Derivation of Theorem 2 for a Two-Dimensional Test

Without loss of generality assume that $|\boldsymbol{a}_j| = 1$ holds for $j = 1, \ldots, k$.

According to Theorem 1, i.e., according to Theorem 6.1 of Hooker et al. (2009), a sufficient condition to ensure a paradoxical effect with respect to the first dimension by a change of the ith item score is that $\boldsymbol{e}_1(A_{-i}^T W A_{-i})^{-1} \boldsymbol{a}_i < 0$ holds for all positive diagonal matrices W. Equivalently, a sufficient condition to ensure the paradoxical effect is that the first component of the solution $\boldsymbol{x}$ to the equation $A_{-i}^T W A_{-i} \boldsymbol{x} = \boldsymbol{a}_i$ is negative for all choices of positive diagonal matrices W. According to Cramer's rule (see e.g., p. 53 in Lax, 2007), the first component may be computed up to a positive scalar $c = (\det(A_{-i}^T W A_{-i}))^{-1}$ as:

$$x_1 = \det \begin{pmatrix} a_{i1} & \sum_{j \neq i} w_j a_{j1} a_{j2} \\ a_{i2} & \sum_{j \neq i} w_j a_{j2}^2 \end{pmatrix} = a_{i1} \sum_{j \neq i} w_j a_{j2}^2 - a_{i2} \sum_{j \neq i} w_j a_{j1} a_{j2}.$$

Due to $a_{i1} < a_{j1}$ and $a_{i2} > a_{j2}$ for all j (here we use the fact that item i is the item with the least relative emphasis on the first dimension and the fact that the test is two-dimensional; the latter implies that item i is also the item with the highest relative loading on the second dimension) it follows that

$$x_1 = \sum_{j \neq i} a_{i1} w_j a_{j2}^2 - \sum_{j \neq i} a_{i2} w_j a_{j1} a_{j2} < \sum_{j \neq i} a_{j1} w_j a_{j2}^2 - \sum_{j \neq i} w_j a_{j1} a_{j2}^2 = 0.$$

Thus, $x_1 < 0$ irrespective of the entries w_l of the positive diagonal matrix W and we may conclude that item i induces a paradoxical movement with respect to the first dimension.

3.4. Extension to Higher-Dimensional Settings: New Insights into the Causation of the Paradox

As a consequence of the preceding discussion in Sects. 3.1 and 3.2, the existence of the paradoxical effect for $p > 2$ dimensional linearly compensatory models could be called into question. That is, the claim that the paradox is omnipresent for this type of models needs to be reinvestigated in the light of the error in the proof. [The error only refers to one specific theorem; it does not call into question the general analysis and framework provided by Hooker et al. (2009)].

In this section, we shall show, however, that the paradoxical effect is also certain to occur in $p > 2$ dimensional models. To achieve this main result, specific items will be singled out which will always guarantee paradoxical scoring effects. Further, the already known condition for the two-dimensional case will be shown to be a special case of the result. As a byproduct, the proof will furnish a rather peculiar additional property of this class of models. We strongly recommend not to skip the proof, as the derivation sheds light on the causation of the paradox.

3.4.1. Proof of the Existence of Paradoxical Effects in Any Linearly Compensatory Model

Proof. We will first assume that each item discrimination vector $\boldsymbol{a}_i$ consists only of positive entries. This will allow for a simplified proof. Later, we will also provide the extension for the case that some entries equal zero.

We denote by f_1 the loglikelihood corresponding to the response pattern $u_1, \ldots u_{k-1}$, i.e., the full loglikelihood excluding the contribution of the kth item (see also Sect. 2). The complete loglikelihood of all responses (including the kth) is denoted as f_2. Thus,

$$f_1(\boldsymbol{\theta}) := \sum_{i \neq k} l_i(\boldsymbol{a}_i^T \boldsymbol{\theta}), \quad f_2(\boldsymbol{\theta}) := \sum_{i=1}^{k} l_i(\boldsymbol{a}_i^T \boldsymbol{\theta}).$$

We will show that the corresponding MLEs cannot be ordered (provided we select an appropriate item k). More specifically, we will argue by contradiction and assume that the MLEs corresponding to the two loglikelihood functions are ordered, i.e., if $\boldsymbol{\theta}$ denotes the MLE for the reduced loglikelihood f_1 and if $\widetilde{\boldsymbol{\theta}}$ denotes the MLE for the full loglikelihood f_2, then $\boldsymbol{\theta} < \widetilde{\boldsymbol{\theta}}$ will be assumed and a contradiction based on this will be derived (here "<" has to be interpreted in the sense of the partial ordering in $\mathbb{R}^p$, i.e., $\theta_i \leq \widetilde{\theta}_i$ for all i and $\theta_j < \widetilde{\theta}_j$ for at least one dimension j).

As the two MLEs are maximizers and as the functions involved are differentiable, each directional derivative must vanish at the MLEs. That is, for each direction $\boldsymbol{v} \in \mathbb{R}^p$ the following equations need to hold [the directional derivative of a function f at the point $\boldsymbol{x}$ in the direction $\boldsymbol{u}$ is abbreviated as $f'(\boldsymbol{x}, \boldsymbol{u})$]:

$$f_1'(\boldsymbol{\theta}, \boldsymbol{v}) = \sum_{i \neq k} l_i'(\boldsymbol{a}_i^T \boldsymbol{\theta}) \boldsymbol{a}_i^T \boldsymbol{v} = 0 \quad \forall \boldsymbol{v}, \tag{1}$$

$$f_2'(\widetilde{\boldsymbol{\theta}}, \boldsymbol{v}) = \sum_{i} l_i'(\boldsymbol{a}_i^T \widetilde{\boldsymbol{\theta}}) \boldsymbol{a}_i^T \boldsymbol{v} = 0 \quad \forall \boldsymbol{v}. \tag{2}$$

The loglikelihood contributions l_i satisfy $l_i'' < 0$ globally by assumption of a linearly compensatory model–see (b) of Sect. 2. Therefore, their derivatives $l_i'(\cdot)$ are strictly decreasing. As all item discrimination parameters are positive, the assumption $\boldsymbol{\theta} < \widetilde{\boldsymbol{\theta}}$ implies $\boldsymbol{a}_i^T \boldsymbol{\theta} < \boldsymbol{a}_i^T \widetilde{\boldsymbol{\theta}}$ for all i.

Setting $w_i := l_i'(\boldsymbol{a}_i^T \boldsymbol{\theta})$ for $i = 1, \ldots k-1$ and $\widetilde{w}_i := l_i'(\boldsymbol{a}_i^T \widetilde{\boldsymbol{\theta}})$ for $i = 1, \ldots k$, it follows from the preceding observations (i.e., from the monotonicity of l_i' and the ordering $\boldsymbol{a}_i^T \boldsymbol{\theta} < \boldsymbol{a}_i^T \widetilde{\boldsymbol{\theta}}$) that $w_i > \widetilde{w}_i$ holds for $i = 1, \ldots k-1$. Subtracting Eqs. (1) from (2) leads to

$$\widetilde{w}_k \boldsymbol{a}_k^T \boldsymbol{v} + \sum_{i \neq k} u_i \boldsymbol{a}_i^T \boldsymbol{v} = 0 \; \forall \boldsymbol{v}, \tag{3}$$

wherein $u_i := \widetilde{w}_i - w_i < 0$ (for $i = 1, \ldots k-1$). Suppose there is a direction[6] $\boldsymbol{v}$ such that $\boldsymbol{a}_k^T \boldsymbol{v} = 0$ and $\boldsymbol{a}_i^T \boldsymbol{v} \leq 0$ holds for all $i = 1, \ldots k-1$. Then, the left-hand side of (3) is positive and a contradiction arises (note that at least one of the inequalities, $\boldsymbol{a}_i^T \boldsymbol{v} \leq 0$, has to be strict, because otherwise all discrimination vectors are contained in the subspace orthogonal to $\boldsymbol{v}$, implying that the matrix A is not of full column rank).

Thus, we have established the following key result:

(C1) If there is an item k and a vector $\boldsymbol{v}$ orthogonal to $\boldsymbol{a}_k$ such that $\boldsymbol{a}_i^T \boldsymbol{v} \leq 0$ holds for all i, then the MLEs of the reduced and the complete loglikelihood cannot be ordered.

Observing that the set $\{\boldsymbol{a} \in \mathbb{R}^p | \boldsymbol{a}^T \boldsymbol{v} = 0\}$ describes a hyperplane through the origin, the above condition can be recast as follows:

(C2) If there is hyperplane (see Fig. 3) through the origin containing an item discrimination vector $\boldsymbol{a}_k$ such that all remaining item discrimination vectors lie in only one of the two associated half spaces (i.e., either $\boldsymbol{a}_i^T \boldsymbol{v} \geq 0$ for all i or $\boldsymbol{a}_i^T \boldsymbol{v} \leq 0$ for all i), then the MLEs of the reduced and the complete loglikelihood cannot be ordered.

---

[6]By definition, a direction $\boldsymbol{v}$ is different from zero as the zero vector has no direction.
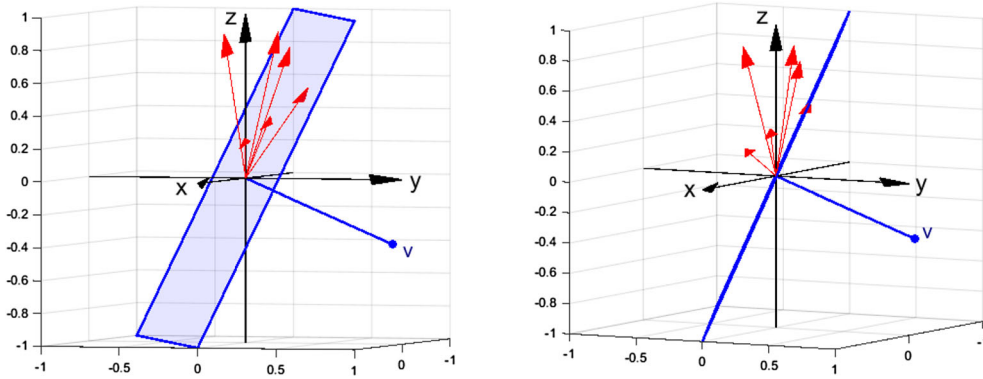
FIGURE 3.

A hyperplane passing through the origin (with normal vector v) such that all discrimination vectors (red) lie in only one of the two associated half space. In this case, changing the item response of the item which is contained in the hyperplane will induce a paradoxical ordering effect (Color figure online).

We can arrive at yet another equivalent geometrical characterization by defining the set of all nonnegative linear combinations of the item discrimination vectors, i.e., $C := \{\sum_i \lambda_i \boldsymbol{a}_i | \lambda_i \geq 0\}$. The latter set is closed under addition and positive scalar multiplications. It is therefore a convex cone—namely the convex cone generated by the item discrimination vectors. The set $C_i := \{\lambda \boldsymbol{a}_i \| \lambda \geq 0\}$ will be called the ray corresponding to the ith item. It is an unbounded half line emanating from the origin in the direction of the item discrimination vector. With this terminology, we can provide a further geometrical description:

(C3) If there is a linear function which achieves its maximum relative to C, the convex cone generated by all item discrimination vectors, at all points of the kth item ray, then the MLEs of the reduced and the complete loglikelihood cannot be ordered.

Of course, the above derivation is based on the existence of such a linear function/hyperplane. The latter is however always guaranteed due to a general result in convex analysis—for a proof of which we refer the reader to Theorem 18.7 of Rockafellar (1970).

We summarize the conclusion of the above derivation as follows:

Theorem 3. *Let a linearly compensatory model with strictly positive item discrimination vectors be given and let $\boldsymbol{a}_k$ be an item discrimination vector such that there is a hyperplane through the origin containing $\boldsymbol{a}_k$ and such that all remaining item discrimination vectors lie in only one of the two associated half spaces. Then the two MLEs corresponding to the reduced loglikelihood and the full loglikelihood are either identical or they cannot be ordered. This holds irrespective of the specific response pattern of the test taker on the kth item.*

Moreover, there are at least p items for which the existence of the hyperplane is guaranteed (see Theorem 18.7 of Rockafellar, 1970).

Note that the independence of the specific response pattern mentioned in Theorem 3 is a remarkable result. It generalizes the results on paradoxical effects, which all seemed to necessitate a certain monotonicity of the loglikelihood contribution of the item whose response is changed (Hooker et al., 2009; Jordan & Spiess, 2012; van der Linden, 2012). That is, in case of a correct response, the function l_k is monotone increasing and this monotonicity property was required in the proof of paradoxical results. However, the above derivation shows that no monotonicity of the contribution of the item is required. The qualitative result on the impossibility of obtaining ordered MLEs holds also for a nonmonotone contribution l_k. This independence result has a further

consequence. It basically states that if we administer an item k with the property stated in C1 (which may be checked solely on the ground of the discrimination parameters!), then we know in advance that the test taker cannot achieve increases on all estimates for his latent abilities, irrespective of the particular response he provides. Each response u_k will result in a different contribution l_{k,u_k} (monotone for the extreme responses in an ordinal IRT model and nonmonotone for intermediate responses)—the qualitative statement on the impossibility of obtaining ordered MLEs nevertheless remains true.

Before moving toward a further extension of the above results, we show that the condition of Corollary 6.1 of Hooker et al. (2009) (rephrased here in Theorem 2) can be derived via the above result, when restricted to the case of a two-dimensional test.

Let a_k be the item with the largest angle to the first coordinate axis. Define $v := (v_1, v_2)^T = (-a_{k,2}, a_{k,1})^T$. Then v is orthogonal to a_k and if a_j is any other item discrimination vector then

$$v^T a_j = -a_{k,2} a_{j,1} + a_{k,1} a_{j,2}. \tag{4}$$

Now according to the angle property (see Theorem 2), the following implication holds:

$$a_{k,1} < a_{j,1} \frac{|a_k|}{|a_j|}.$$

Because of the two-dimensional setting it likewise follows that:

$$a_{k,2} > a_{j,2} \frac{|a_k|}{|a_j|}.$$

Using these two inequalities in (4) leads to:

$$v^T a_j = -a_{k,2} a_{j,1} + a_{k,1} a_{j,2} < -a_{j,2} \frac{|a_k|}{|a_j|} a_{j,1} + a_{j,1} \frac{|a_k|}{|a_j|} a_{j,2} = 0.$$

Thus, the condition on the angle implies that there is a hyperplane through the origin with normal vector $(-a_{k,2}, a_{k,1})^T$ satisfying the condition (C1) which in turn implies the existence of a paradoxical effect according to Theorem 3.

3.4.2. Extending the Results

We first sketch how the assumption of strictly positive item discrimination vectors can be relaxed:

Let the item discrimination vectors merely satisfy nonnegativity and let v and a_k be as described in (C1). If there is a single item j with strict positive item discrimination whose discrimination vector is not contained in the hyperplane, i.e., $a_j^T v < 0$, then an inspection of the proof shows that the conclusion of Theorem 3 still holds.

Following the reasoning in the proof, a number of further extensions are possible—some of them have quite interesting consequences. We first list these possible extensions in increasing order of importance and comment on the meaning of the most important one in the discussion.

(1) The assumption of twice continuously differentiable functions l_i satisfying $l_i'' < 0$ can be replaced by the weaker assumption of differentiable functions l_i with strictly monotone decreasing derivatives l_i'.

(2) The assumption of the uniqueness of the MLEs can be dropped. Of course, this does not make sense in the light of the strict concavity of the loglikelihoods, but it broadens the scope with respect to the extensions (4) and (5) listed below.

(3) Using the implicit function theorem, the existence of paradoxical scoring effects can also be deduced for an item not satisfying the condition $C1$—as long as the item discrimination vector is within a neighborhood of some item discrimination vector satisfying the condition.

(4) If a_k has the properties stated in (C1), then the shape of the loglikelihood contribution l_k does not matter (provided l_k is differentiable and that the existence of the MLE is still guaranteed) with respect to the conclusion that the two MLEs cannot be strictly ordered.

(5) In fact, (4) generalizes further: let v be as stated in (C1) and let I be the index set of all items whose discrimination vectors lie in the hyperplane through the origin with normal vector v. Then, the shape of all the functions l_i with $i \in I$ does not matter with respect to the conclusion that the two MLEs cannot be strictly ordered.

4. Discussion

The analysis of paradoxical scoring effects, wherein correct answers are penalized and incorrect answers are rewarded, is an important part in terms of test fairness and also for further understanding of the underlying model based estimates. We have shown in this paper that the geometry of the item discrimination vectors—more specifically, the geometry of the item rays pointing in the direction of the item discrimination vectors—is crucial to the understanding and interpretation of ordering properties of MLEs in linearly compensatory item response models. Any item with a discrimination vector satisfying a certain "boundary" condition (namely the existence of a supporting hyperplane containing the item discrimination vector) is incompatible with the notion of ordering. The MLE of the response pattern excluding the item will never be lower (or higher) on all coordinates. Of course, this does not imply that the remaining items, i.e., items with discrimination vectors not satisfying the boundary condition, are free of paradoxical scoring—as the condition (C1) is sufficient, but not necessary for a paradoxical scoring effect.

Moreover, we have derived a quite bizarre extension of the paradoxical effect: if we are given an item with an item discrimination vector satisfying the boundary condition, then we know that the test taker cannot achieve a higher estimate on all dimensions—irrespective of his response pattern and, in addition, irrespective of the specific model which is used for that item (as long as it is a function of $a^T \theta$). This emphasizes the incompatibility of the class of linearly compensatory models with common notions of test fairness (or more generally: of models using linear composite terms $a^T \theta$ in conjunction with "primarily" log concave contributions).

Importantly, our method does not lead to a dimension-specific conclusion. That is, we know that the MLEs cannot be ordered—yet a precise statement regarding the components is lacking. Therefore, with the exception of the two-dimensional case, there remains the open question as to whether each component is affected by the paradox. Or stated differently: Could it be possible to construct a three- or higher-dimensional test, wherein one latent dimension is free of paradoxical scoring effects? Although intuition would suggest that this probably is not possible, supplying this intuitive notion with a rigorous proof is surprisingly difficult (we expect either a linear algebra approach based on Theorem 1 or tools from constrained optimization such as the conjugacy operation to be useful in providing the ultimate answer). Therefore, this topic remains a subject for further research.

More generally, it is difficult to tell which item rewards correct answers on which dimension. For example, the counterexample presented in Sect. 3.1 shows that the first item has equal loadings on the first and second dimension—yet the sign of the scoring on these dimensions differs, i.e., the

second dimension is affected by paradoxical scoring, whereas the first dimension is unaffected. Moreover, if an item discriminates highly along one dimension (see the rule which was rephrased in the introduction), one expects the item score to have a large impact on the scoring on that dimension—yet the third item provides an example wherein the scoring of the dimension is unaffected, despite the high (relative) loading of 4 on that dimension. These points highlight the counterintuitive relationship between the item discrimination vectors and the estimate(s) of the latent abilities. (Note that these observations provide a critique for the usage of intuitive heuristics with respect to the scoring of the items but that they do not have any bearing on the interpretation of item discrimination vectors in terms of slopes/odds-ratios.)

Finally, it should also be noted that—though paradoxical effects are certain to occur—there is up to date a general lack of studies which examine the size of these effects in practical applications. To the authors' knowledge, only the study of Finkelman, Hooker and Wang (2010) provided a systematic analysis of the magnitude of paradoxical effects in a real data framework. Finkelman, Hooker and Wang used a large data set[7] from an English Language Learners program—a test which is scored two-dimensional with one reading/writing dimension and one listening dimension—to calculate the percentage of test takers which could have received a more favorable classification on the listening dimension by changing one or more correct answers into incorrect answers. Depending on the threshold that is used to define a positive classification, the proportion varied between 4 and 20% (see Figure 7 in Finkelman, Hooker & Wang, 2010). This result therefore clearly demonstrates that the paradoxical effect is not a theoretical artifact. Opinions differ, however, on how to cope with this effect (Hooker et al., 2009; van der Linden, 2012; van Rijn & Rijmen, 2012; Reckase & Luo, 2015). On the one hand, the underlying statistical estimator is optimal and one could make a valid point that the paradoxical effect is an inherent part of this optimality. Therefore, it could also be seen as a byproduct of statistically precise estimation (Reckase & Luo, 2015). If that were the case, then any attempts to "fix" the effect would also tend to remove favorable statistical properties (of course this latter point would have serious implications with respect to the usage of ordinary sumscores in practice). On the other hand, the presence of this effect poses a serious threat to common notions of test fairness—especially in high-stake testing situations (Hooker et al, 2009). That is, despite of any favorable statistical properties, the results of the paradoxical effect can be a challenge in terms of social acceptability (van Rijn & Rijmen, 2012). The general notion that correct answers should be rewarded and incorrect answers should be penalized is hardwired into our common understanding of test fairness. Therefore, we think that the phenomenon of paradoxical scoring effects needs to be addressed beyond the perspective which views the paradox as "just" an instance of a statistically reasonable effect.

Appendix A: Proof of Theorem 1

Define a function $f : \mathbb{R}^p \times \mathbb{R} \mapsto \mathbb{R}^p$ of $\boldsymbol{\theta}$ and a perturbation parameter $\lambda \in \mathbb{R}$ as follows:

$$f(\boldsymbol{\theta}, \lambda) := \sum_{i \neq k} l_i'(\boldsymbol{a}_i^T \boldsymbol{\theta}) \boldsymbol{a}_i^T + \lambda l_k'(\boldsymbol{a}_k^T \boldsymbol{\theta}) \boldsymbol{a}_k^T.$$

Note that the solution $\boldsymbol{\theta}$ satisfying $f(\boldsymbol{\theta}, \lambda = 1) = \mathbf{0}$ can be equated with the MLE corresponding to the full loglikelihood, whereas the solution $\boldsymbol{\theta}$ satisfying $f(\boldsymbol{\theta}, \lambda = 0) = \mathbf{0}$ can be equated with the MLE corresponding to the reduced loglikelihood. An application of the implicit function

---

[7]The underlying modeling framework differed from the setup of this paper in two ways: firstly, Bayesian estimates were computed rather than MLEs and secondly, the model underlying the two-dimensional test was not logconcave. We report the Bayesian estimates corresponding to an independence prior which resembles the MLE-framework more closely than highly dependent priors.

theorem (see e.g., Dontchev & Rockafellar, 2009) provides the derivative of the solutions as a function of the perturbation parameter λ:

$$s'(\lambda) = -\left(\sum_{i \neq k} \boldsymbol{a}_i l_i''(\boldsymbol{a}_i^T \boldsymbol{\theta})\boldsymbol{a}_i^T + \lambda \boldsymbol{a}_k l_k''(\boldsymbol{a}_k^T \boldsymbol{\theta})\boldsymbol{a}_k^T\right)^{-1} l_k'(\boldsymbol{a}_k^T \boldsymbol{\theta})\boldsymbol{a}_k.$$

Defining weights $w_i := l_i''(\boldsymbol{a}_i^T \boldsymbol{\theta})(i \neq k)$, $w_k = \lambda l_k''(\boldsymbol{a}_k^T \boldsymbol{\theta})$—which are negative due to assumption (b)—the above expression can also be written as

$$s'(\lambda) = -l_k'(\boldsymbol{a}_k^T \boldsymbol{\theta})\left(A^T W A\right)^{-1} \boldsymbol{a}_k,$$

with W a diagonal matrix containing the weights. If the loglikelihood contribution of the kth item corresponds to a correct response, then l_k is strictly increasing. Thus $l_k' > 0$ holds and $s'(\lambda)$ will be a decreasing function in the first component if

$$\boldsymbol{e}_1^T \left(A^T W A\right)^{-1} \boldsymbol{a}_k^T > 0$$

holds for all negative diagonal matrices W—which is almost the condition as cited in Theorem 1. Using the Sherman–Morrison formula (see e.g., p. 301 of Puntanen, Styan & Isotalo, 2011), it can be shown that the sign of $\boldsymbol{e}_1^T \left(A^T W A\right)^{-1} \boldsymbol{a}_k^T$ is identical to the sign of the expression as stated in Theorem 6.1 of Hooker et al. (2009).

Appendix B: A Counterexample to Theorem 2 for a Binary MIRT Model

The counterexample (Sect. 3.1) to Theorem 2 was based on a model—the classical factor analysis model—which is not a proper IRT model and which does not fall within the original framework addressed by Hooker et al. (2009) when deriving paradoxical results (although the derivations can be transferred to this setting). The purpose of this appendix is to provide R-Code demonstrating that the rule given in Theorem 2 is also false within a binary MIRT framework. To achieve this, we simulate response data from a M2PL model of test length $k = 30$. The test length of $k = 30$ is chosen to ensure (with high probability) the existence of the MLE. Although the counterexample uses a seed for reproducibility, we encourage the reader to deviate from this and to simulate other test structures and response data according to the program below in order to convince himself that our particular counterexample does not represent a special case but generalizes to other test structures.

```
set.seed(53678)
#Simulate three-dimensional test structure of test length k=30
#Each item discrimination vector is positive (and of equal length)
sim_matrix <- function() {
    A <- matrix(runif(30 * 3), 30, 3, byrow = TRUE)
    A / sqrt(rowSums(A^2))
}

#Simulate an arbitrary response pattern
```

```
#For the purpose of demonstrating paradoxical results,
#we just need any pattern with a finite MLE. It is not necessary
#to generate a response pattern according to the IRT model
sim_response <- function()
{y <- as.numeric(runif(30)<0.5)
}


A <- sim_matrix()
y <- sim_response()

#Find the item with largest angle to the first coordinate axis
index <- which.min(A[,1])

#Generate a new response pattern identical to the
#generated response - except for the
#coordinate on the item with largest angle.
y[index]
y2 <- y
y2[index] <- 1-y[index]

#Compare the MLEs of the first components
glm(y~A-1,family=binomial(link="logit"))$coef
glm(y2~A-1,family=binomial(link="logit"))$coef
```

References

Dontchev, A. L., & Rockafellar, R. T. (2009). *Implicit functions and solution mappings: A view from variational analysis*. Berlin: Springer.

Finkelman, M. D., Hooker, G., & Wang, Z. (2010). Prevalence and magnitude of paradoxical results in multidimensional item response theory. *Journal of Educational and Behavioral Statistics*, *35*(6), 744–761.

Hooker, G. (2010). On separable tests, correlated priors, and paradoxical results in multidimensional item response theory. *Psychometrika*, *75*(4), 694–707.

Hooker, G., Finkelman, M., & Schwartzman, A. (2009). Paradoxical results in multidimensional item response theory. *Psychometrika*, *74*(3), 419–442.

Jordan, P., & Spiess, M. (2012). Generalizations of paradoxical results in multidimensional item response theory. *Psychometrika*, *77*, 127–152.

Karlin, S., & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities II. Multivariate reverse rule distributions. *Journal of Multivariate Analysis*, *10*(4), 499–516.

Lax, P. (2007). *Linear algebra*. New York: Wiley.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

Puntanen, S., Styan, G. P., & Isotalo, J. (2011). *Matrix tricks for linear statistical models: Our personal top twenty*. New York: Springer.

Reckase, M. D., & Luo, X. (2015). A paradox by another name is good estimation. In R. Millsap, D. Bolt, L. van der Ark, & W. C. Wang (Eds.), *Quantitative Psychology Research. Springer Proceedings in Mathematics & Statistics* (Vol. 89). Cham: Springer.

Reckase, M. (2009). *Multidimensional item response theory*. New York: Springer.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press.

Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, *39*(1), 111–121.

van der Linden, W. J. (2012). On compensation in multidimensional response modeling. *Psychometrika*, *77*, 21–30.

van Rijn, P. W., & Rijmen, F. (2012). *A note on explaining away and paradoxical results in multidimensional item response theory (ETS No. RR-12-13)*. Princeton: Educational Testing Service.