SOLUTIONS OF ORTHOGONAL PROCRUSTES PROBLEMS UNDER PARTIALLY KNOWN PRIOR

JUN XU

DEPARTMENT OF COMPUTING, THE HONG KONG POLYTECHNIC UNIVERSITY

This aricle is based on the Presidential Address William Stout gave on June 23, 2002 at the 67th Annual Meeting of the Psychometric Society held in Chapel Hill, North Carolina. —Editor

I wish to especially thank Sarah Hartz and Louis Roussos for their suggestions that helped shape this paper. I wish to thank all my former Ph.D. students: Without their contributions, the content of this paper would have been vastly different and much less interesting!

Requests for reprints should be sent to William Stout, Department of Statistics, University of Illinois, 725 S. Wright Street, Champaign IL 61820. E-Mail: stout@stat.uiuc.edu

Dedication: I want to dedicate this paper to my wife, Barbara Meihoefer, who was lost to illness in this year of my presidency. For, in addition to all the wonderful things she meant to me personally and the enormous support she gave concerning my career, she truly enjoyed and greatly appreciated my psychometric colleagues and indeed found psychometrics an important and fascinating intellectual endeavor, in particular finding the skills diagnosis area exciting and important: She often took time from her career as a business manager and entrepreneur to attend psychometric meetings with me and to discuss research projects with my colleagues and me. She would have enjoyed this paper.—William Stout

Abstract

The orthogonal Procrustes problem aims to find an orthogonal matrix which transforms one given matrix into another by minimizing their Frobenius matrix norm. This problem can be applied in applications such as permutation theory, machine learning, and camera calibration, *etc*. In real cases, the permutation matrix may have been partially known, the dictionaries can be partially learned from external data, and the calibration of camera should be done under some fixed priors. This prior information makes the original orthogonal Procrustes problem more difficult. In this paper, we consider the solution of this problem under partially known priors, which includes the original orthogonal Procrustes problem as a special case with no such prior.

Key words: orthogonal Procrustes problem, pratially known priors

1. Introduction

The classical orthogonal Procrustes problem has been applied in factor analysis and statistics, machine learning, computer vision, optical imaging, and robotics.

2. Definition of the Probelm and Solution

Let $\mathbf{A}, \mathbf{B} \in \mathcal{R}^{n\times m}$ be two given data matrices. Define $\mathbf{X} \in \mathcal{R}^{n\times p}$ and $\mathbf{P} \in \mathcal{R}^{n\times q}$ where $p + q = n$. $\mathbf{X}$ is the partially known prior which could be used to guide the solutions of $\mathbf{P}$. We formulate the orthogonal Procrustes problem with partially known priors as:

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{D}} \|\mathbf{B} - [\mathbf{X}\ \mathbf{P}]\mathbf{A}\|_F^2 \quad s.t. \quad \mathbf{P}^\top\mathbf{P} = \mathbf{I}_{q\times q}, \mathbf{X}^\top\mathbf{P} = \mathbf{0}_{p\times q}. \tag{1}$$

In fact, as have been proofed, if the matrix $\mathbf{BA}^\top$ has no zero singular value, then the solution of $\hat{\mathbf{P}} = \mathbf{UV}^\top$ is unique and we do not need any preceding results.

We crop the matrix $\mathbf{A}$ into two parts: $\mathbf{A}_X \in \mathcal{R}^{p\times m}$ and $\mathbf{A}_P \in \mathcal{R}^{q\times m}$ to interact with $\mathbf{X}$ and $\mathbf{P}$, respectively. Then we have

$$\begin{aligned} &\|\mathbf{B} - [\mathbf{X}\ \mathbf{P}]\mathbf{A}\|_F^2 = \|\mathbf{B} - [\mathbf{X}\ \mathbf{P}][\mathbf{A}_X^\top\ \mathbf{A}_P^\top]^\top\|_F^2 = \|\mathbf{B} - [\mathbf{X}\ \mathbf{P}][\mathbf{A}_X^\top\ \mathbf{A}_P^\top]^\top\|_F^2 \\ &= \|\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top - \mathbf{P}\mathbf{A}_P^\top]\|_F^2 = \|\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top - \mathbf{P}\mathbf{A}_P^\top]\|_F^2 \end{aligned} \tag{2}$$

The $\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$ is a known data matrix and we replace it with $\mathbf{B}^* = \mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$. In the following Results 1, we remove the notation $*$ and use $\mathbf{B}$ as the finally known data matrix.

Results 1: The solution of

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{D}} \|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 \quad s.t. \quad \mathbf{P}^\top\mathbf{P} = \mathbf{I}_{q\times q}, \mathbf{X}^\top\mathbf{P} = \mathbf{0}_{p\times q} \tag{3}$$

is $\hat{\mathbf{P}} = \mathbf{UV}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are the orthogonal matrices obtained by perfroming economy (aka. reduced) SVD:

$$(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \tag{4}$$

proof:

2.1. Unidimensionality from the Weak LI Conditional Covariance Perspective

2.2. Foundational Issues Facilitated by Infinite Test Length Unidimensional MLI1 Modeling

2.3. Interpreting Conditional Covariances Geometrically to Assess Latent Multidimensional Structure

2.4. NIRT-Based Statistical Procedures, Emphasizing Conditional Covariances

FIGURE 1.
Projection of item discrimination vectors onto V_{θ_T} hyperplance for a six item three-dimensional approximate sample structure.

3. Test Fairness

3.1. Multidimensional Model for DIF (MMD)

3.2. Model-Based Parameterization of the amount of DIF in Various Settings

3.3. MMD- Inspired DIF Statistical Procedures

FIGURE 2.
Comparison of Θ_F and Θ_R distribution with $\Theta_F|X_V = k$ and $\Theta_R|X_V = k$ distributions.

3.4. Implementation of DIF/DBF Procedures

FIGURE 3.
Item discrimination vectors of a 22 item validity sector.

FIGURE 4.
Panel index versus bundle DBF $\hat{\beta}$/item.

4. Formative Assessment Skills Diagnosis: A New Test Paradigm

4.1. A Brief Survey of Psychometric Skills Diagnostic Models

4.2. The Unified Model and Generalizations Making it Useful

4.3. Application of the Unified Model to PSAT Data

4.4. Skills Diagnosis: The New Paradigm?

5. Dimensionality, Equity, and Diagnostic Software

6. Concluding Remarks

FIGURE 5.
North Carolina End-of-Grade Math Skills Test Subscores.

FIGURE 6.
PSAT Score Report *Plus* Skills Mastery Reporting.

References

Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, *29*, 67–91.

Angoff, W.H. (1993). Perspectives on differential item functioning methodology. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning*(pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bolt, D., Froelich, A.G., Habing, B., Hartz, S., Roussos, L., & Stout, W. (in press). *An applied and foundational research project addressing DIF, impact, and equity: With applications to ETS test development* (ETS Technical Report). Princeton, NJ: ETS.

Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: an adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, *33*, 333–353

Chang, H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika*, *58*, 37–52.

DiBello, L., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. Chipman, & R. Brennen (Eds.), *Cognitively diagnostic assessment* (pp. 361–389). Hillsdale, NJ: Earlbaum.

Doignon, J.-P., & Falmagne, J.-C. (in press). *Knowledge spaces*. Berlin Springer-Verlag.

Dorans, N.J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355–368.

Douglas, J. (1997). Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychometrika*, *62*, 7–28.

Douglas, J.A. (2001). Asymptotic identifiability of nonparametric item response models. *Psychometrika*, *66*, 531–540.

Douglas J.A., & Cohen A. (2001). Nonparametric ICC estimation to assess fit of parametric models. *Applied Psychological Measurement*, *25*, 234–243.

Douglas, J., Kim, H.R., Habing, B., & Gao, F. (1998) Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, *23*, 129–151.

Douglas, J., Roussos, L., & Stout, W., (1996). Item bundle DIF hypothesis testing: Identifying suspect bundles and assessing their DIF. *Journal of Educational Measurement*, *33*, 465–484.

Douglas, J., Stout, W., & DiBello, L. (1996). A kernel smoothed version of SIBTEST with applications to local DIF inference and unction estimation. *Journal of Educational and Behavioral Statistics*, *21*, 333–363.

Ellis, J.L., & Junker, B.W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika*, *62*, 495–524.

Embretson (Whitely), S.E. (1980). Multicomponent latent trait models for ability tests *Psychometrika*, *45*, 479–494.

Embretson, S.E. (1984). A general latent trait model for response processes. *Psychometrika*, *49*, 175–186.

Embretson, S. E. (Ed.). (1985), *Test design: Developments in psychology and psychometrics* (pp. 195–218, chap. 7). Orlando, FL: Academic Press.

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*, 359–374.

Froelich, A.G., & Habing, B. (2002, July). A study of methods for selecting the AT subtest in the DIMTEST procedure. Paper presented at the 2002 Annual Meeting of the Psychometrika Society, University of North Carolina at Chapel Hill.

Gierl, M.J., Bisanz, J., Bisanz, G., Boughton, K., & Khaliq, S. (2001). Illustrating the utility of differential bundle functioning analyses to identify and interpret group differences on achievement tests. *Educational Measurement: Issues and Practice*, *20*, 26–36.

Gierl, M.J., & Khaliq, S.N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests. *Journal of Educational Measurement*, *38*, 164–187.

Gierl, M.J., Bisanz, J., Bisanz, G.L., & Boughton, K.A. (2002, April). Identifying content and cognitive skills that produce gender differences in mathematics: A demonstration of the DIF analysis framework. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Haberman, S.J (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, *5*, 815–841.

Habing, B. (2001). Nonparametric regression and the parametric bootstrap for local dependence assessment. *Applied Psychological Measurement*, *25*, 221–233.

Haertel, E. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301–321.

Hartz, S.M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: blending theory with practicality*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign, Department of Statistics.

Holland, P.W. (1990a). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577–601.

Holland, P.W. (1990b). The Dutch identity: a new tool for the study of item response models. *Psychometrika*, *55*, 5–18.

Holland, P.W., & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, *14*, 1523–1543.

Holland, W.P., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Earlbaum Associates.

Jiang, H., & Stout, W. (1998). Improved Type I error control and reduced estimation bias for DIF detection using SIBTEST. *Journal of Educational and Behavioral Statistics*, *23*, 291–322.

Junker, B.W. (1993). Conditional association, essential independence, and monotone unidimensional latent variable models. *Annals of Statistics*, *21*, 1359–1378.

Junker, B.W. (1999). *Some statistical models and computational methods that may be useful for cognitively-relevant assessment*. Prepared for the National Research Council Committee on the Foundations of Assessment. Retrieved April 2, 2001, from http://www.stat.cmu.edu/∼brian/nrc/cfa/

Junker, B.W., & Ellis, J.L. (1998). A characterization of monotone unidimensional latent variable models. *Annals of Statistics*, *25*(3), 1327–1343.

Junker, B. W. & Sijtsma, K. (2001). Nonparametric item response theory in action: an overview of the special issue. *Applied Psychological Measurement*, *25*, 211–220.

Koedinger, K.R., & MacLaren, B.A. (2002). Developing a pedagogical domain theory of early algebra problem solving (CMU-HCII Tech. Rep. 02–100). Pittsburgh, PA: Carnegie Mellon University, School of Computer Science.

Li, H. & Stout, W. (1996). A new procedure for detecting crossing DIF. *Psychometrika*, *61*, 647–677.

Kok, F. (1988). Item bias and test multidimensionality. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent models* (pp. 263–275). New York, NY: Plenum Press.

Linn, R.L. (1993). The use of differential item functioning statistics: A discussion of current practice and future implications. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 349–364). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F.M. (1980) *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates, Hinsdale, NJ.

McDonald, R.P. (1994). Testing for approximate dimensionality. In D. Laveault, B.D. Zumbo, M.E. Gessaroli, & M.W. Boss (Eds.), *Modern theories of measurement: Problems and issues* (pp. 63–86). Ottawa, Canada: University of Ottawa.

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*, 523–547.

Mislevy, R.J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*, 439–483.

Mislevy, R.J. Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K.B. Laskey & H. Prade (Eds.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp. 437–446). San Francisco, CA: Morgan Kaufmann.

Mislevy, R., Steinberg, L. & Almond, R. (in press). On the structure of educational assessments. *Measurement: Interdisciplinary research and perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mokken, R.J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton.

Molenaar, I.W., & Sijtsma, K. (2000). *User's manual MSP5 for Windows: A program for Mokken Scale Analysis for Polytomous Items. Version 5.0* [Software manual]. Groningen: ProGAMMA.

Nandakumar, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement*, *30*, 293–311.

Nandakumar, R., & Roussos, L. (in press). Evaluation of CATSIB procedure in pretest setting. *Journal of Educational and Behavioral Statistics*.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics*, *18*, 41–68.

O'Neill, K.A., & McPeek, W.M. (1993). Item and test characteristics that are associated with differential item functioning. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 255–276). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pellegrino, J.W., Chudowski, N., & Glaser, R (Eds.). (2001). *Knowing what students know: The science and design of educational assessment* (chap. 4, pp. 111–172) Washington, DC: National Academy Press.

Philipp, W. & Stout, W. (1975). *Almost sure convergence principles for sums of dependent random variables* (American Mathematical Society Memoir No. 161). Providence, RI: American Mathematical Society.

Ramsay, J.O. (2000). TESTGRAF: *A program for the graphical analysis of multiple choice test and questionnaire data* (TESTGRAF user's guide for TESTGRAF98 software). Montreal, Quebec: Author. Versions available for Windows®, DOS, and Unix. The Windows® version was retreived November 11, 2002 from ftp://ego.psych.mcgill.ca/pub/ramsay/testgraf/TestGraf98.wpd

Ramsey, P.A. (1993). Sensitivity review: the ETS experience as a case study. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 367–388). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rossi, N., Wang, W. & Ramsay, J.O. (in press). Nonparametric item response function estimates with the EM algorithm. *Journal of Educational and Behavioral Statistics*.

Roussos, L., & Stout, W. (1996a). DIF from the multidimensional perspective. *Applied Psychological Measurement*, *20*, 335–371.

Roussos, L., & Stout, W. (1996b). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type 1 error performance. *Journal of Education Measurement*, *33*, 215–230.

Roussos, L.A., Stout, W.F., & Marden, J. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, *35*, 1–30.

Roussos, L.A., Schnipke, D.A., & Pashley, P.J. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Educational and Behavioral Statistics*, *24*, 293–322.

Shealy, R.T. (1989). *An item response theory-based statistical procedure for detecting concurrent internal bias in ability tests*. Unpublished doctoral dissertation, Department of Statistics, University of Illinois, Urbana-Champaign.

Shealy, R.,& Stout, W. (1993a). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, *58*, 159–194.

Shealy, R., & Stout, W. (1993b). An item response theory model for test bias and differential test functioning. In P. Holland & H. Wainer (Eds.), *Differential item functioning*(pp. 197–240). Hillsdale, NJ: Lawrence Erlbaum.

Sijtsma, K. (1998). Methodology review: nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement*, *22*, 3–32.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*, 589–617.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293–325.

Stout, W., Froelich, A.G., & Gao, F. (2001). Using resampling to produce an improved DIMTEST procedure. In A. Boomsma, M.A.J. van Duijn, T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 357–376). New York, NY: Springer-Verlag.

Stout, W., Habing, B., Douglas, J., Kim, H.R., Roussos, L., & Zhang, J. (1996). Conditional covariance based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, *20*, 331–354.

Stout, W., Li, H., Nandakumar, R., & Bolt, D. (1997). MULTISIB—A procedure to investigate DIF when a test is intentionally multidimensional. *Applied Psychological Measurement*, *21*, 195–213.

Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, *48*, 191–199.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glazer, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. Nichols, S. Chipman, & R. Brennen (Eds.), *Cognitively diagnostic assessment*. Hillsdale, NJ: Earlbaum. 327–359.

Thissen, D., & Wainer, H. (2001). *Test scoring*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trachtenberg, F., & He, X. (2002). One-step joint maximum likelihood estimation for item response theory models. Submitted for publication.

Tucker, L.R., Koopman, R.F., & Linn, R.L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421–459.

Wainer, H., & Braun, H.I. (1988). *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates. Zhang, J., & Stout, W. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64, 129–152.

Whitely, S.E. (1980). (See Embretson, 1980)

Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*, 213–249.