Simo Puntanen
George P. H. Styan
Jarkko Isotalo

Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty

SPIN: 12271245

– Monograph –

June 20, 2011

Springer

Contents

Introduction 1
Opening Figures 1
Preliminaries to Linear Algebra and Statistics 3
Preliminaries to Random Vectors and Data Matrices 16
Preliminaries to Linear Models 28
Exercises 46
Opening Figures: Continued—see page 2 56

1 Easy Column Space Tricks 57
1.1 Equality $\mathscr{C}(\mathbf{X}'\mathbf{V}\mathbf{X}) = \mathscr{C}(\mathbf{X}'\mathbf{V})$ when $\mathbf{V}$ Is Nonnegative Definite 59
1.2 Estimability in a Simple ANOVA 60
1.3 Column Space of the Covariance Matrix 62
1.4 Exercises 66

2 Easy Projector Tricks 71
2.1 Minimal Angle 77
2.2 Minimizing $\mathrm{var}_{\mathrm{d}}(\mathbf{y} - \mathbf{X}_0\mathbf{b})$ 79
2.3 Minimizing $(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})$ 80
2.4 Orthogonal Projector with Respect to $\mathbf{V}$ 81
2.5 Generalized Orthogonal Projector 85
2.6 Exercises 88

3 Easy Correlation Tricks 91
3.1 Orthogonality and Uncorrelatedness 92
3.2 On the Role of the Vector $\mathbf{1}$ 95
3.3 Partial Correlation Geometrically 101
3.4 Exercises 102

4 Generalized Inverses in a Nutshell 105
4.1 Generalized Inverse & the Singular Value Decomposition . . . 109

4.2 Generalized Inverse & Oblique Projector . . . 112
4.3 Generalized Inverse & Solutions to a Linear Equation . . . 114
4.4 A Simple Example . . . 118
4.5 Exercises . . . 119

5 Rank of the Partitioned Matrix and the Matrix Product . . . 121
5.1 Decomposition $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp})$ . . . 123
5.2 Consistency Condition $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$ . . . 125
5.3 Rank of a Partitioned Nonnegative Definite Matrix . . . 126
5.4 Equality of $\mathscr{C}(\mathbf{AF})$ and $\mathscr{C}(\mathbf{AG})$ . . . 126
5.5 Rank of the Model Matrix . . . 127
5.6 The Rank of the Extended Model Matrix . . . 130
5.7 Matrix Volume and the Sample Generalized Variance . . . 131
5.8 Properties of $\mathbf{X}'\mathbf{V}\mathbf{X}$ . . . 132
5.9 Number of Unit Eigenvalues of $\mathbf{P}_{\mathbf{A}}\mathbf{P}_{\mathbf{B}}$ . . . 133
5.10 Properties of $\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{V}\mathbf{X}$ . . . 134
5.11 Intersection $\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B})$ . . . 137
5.12 Solution (in $\mathbf{Y}$) to $\mathbf{Y}(\mathbf{A} : \mathbf{B}) = (\mathbf{0} : \mathbf{B})$ . . . 139
5.13 Some Properties of $\mathscr{C}(\mathbf{X})^{\perp}_{\mathbf{V}^{-1}}$ . . . 140
5.14 Exercises . . . 143

6 Rank Cancellation Rule . . . 145
6.1 Simple Properties of the Hat Matrix $\mathbf{H}$ . . . 146
6.2 Properties of $\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}$ . . . 147

7 Sum of Orthogonal Projectors . . . 151

8 A Decomposition of the Orthogonal Projector . . . 155
8.1 Mahalanobis Distance & the Hat Matrix $\mathbf{H}$ . . . 157
8.2 Subvector of $\hat{\boldsymbol{\beta}}$ . . . 158
8.3 A Single Regression Coefficient . . . 161
8.4 The Reduced Model & the Frisch–Waugh–Lovell Theorem . . . 163
8.5 The Subvectors $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_{\mathbf{x}}$ . . . 166
8.6 The Decomposition SST = SSR + SSE . . . 168
8.7 Representations of R^2 . . . 171
8.8 R as a Maximal Correlation . . . 172
8.9 Testing a Linear Hypothesis . . . 175
8.10 A Property of the Partial Correlation . . . 178
8.11 Another Property of the Partial Correlation . . . 179
8.12 Deleting an Observation: $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$ . . . 180
8.13 Geometrical Illustrations . . . 183
8.14 Exercises . . . 184

9 Minimizing cov(y − Fx) 191
9.1 Best Linear Predictor 194
9.2 What Is Regression? 198
9.3 Throwing Two Dice 201
9.4 Predictive Approach to Principal Component Analysis 203
9.5 A Property of the Prediction Errors 207
9.6 Prediction in the Autocorrelation Structure 208
9.7 Minimizing cov(**Hy** − **FMy**) 210
9.8 Exercises 211

10 BLUE 215
10.1 BLUE & Pandora's Box when $\mathbf{X} = \mathbf{1}_n$ 222
10.2 OLSE vs BLUE when $\mathbf{X} = \mathbf{1}_n$ 224
10.3 OLSE vs BLUE under the Intraclass Correlation Structure . . 225
10.4 General Expressions for the BLUE 228
10.5 Multivariate Linear Model 229
10.6 Random Sample without Replacement 234
10.7 Efficiency of the OLSE 234
10.8 Efficiency and Canonical Correlations 241
10.9 Best Linear Unbiased Predictor 245
10.10 Examples on the BLUPs 250
10.11 Mixed Linear Models 253
10.12 Linear Sufficiency 257
10.13 Admissibility 259
10.14 Exercises 260

11 General Solution to AYB = C 267
11.1 Equality of the BLUEs of $\mathbf{X}\boldsymbol{\beta}$ under Two Models 269
11.2 Model $\mathscr{M}_I$ vs Model $\mathscr{M}_V$ 272
11.3 Stochastic Restrictions and the BLUPs 272
11.4 Two Mixed Models 276
11.5 Two Models with New Unobserved Future Observations 278
11.6 Exercises 280

12 Invariance with Respect to the Choice of Generalized Inverse 283
12.1 Invariance of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ 284
12.2 Estimability and Invariance 284
12.3 Properties of $\mathbf{X}'\mathbf{W}^{-}\mathbf{X}$ 286
12.4 Exercises 288

13 Block-Diagonalization and the Schur Complement 291
13.1 "Inverting" a Partitioned nnd Matrix—Schur Complement . . 294
13.2 Inverting $\mathbf{X}'\mathbf{X}$ 295
13.3 Determinants, Ranks and Schur Complements 297
13.4 Inverting a Sum of Matrices 300

13.5 Generating Observations with a Given Covariance Matrix . . . 301
13.6 Exercises . . . 303

14 Nonnegative Definiteness of a Partitioned Matrix . . . 305
14.1 When Is a "Correlation Matrix" a Correlation Matrix? . . . 307
14.2 $\mathbf{A} \leq_{\mathsf{L}} \mathbf{B} \iff \mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{B})$ and $\mathrm{ch}_1(\mathbf{A}\mathbf{B}^{+}) \leq 1$ . . . 311
14.3 BLUE's Covariance Matrix as a Shorted Matrix . . . 312
14.4 Exercises . . . 314

15 The Matrix $\dot{\mathbf{M}}$ . . . 317
15.1 Representations for the BLUE . . . 324
15.2 The BLUE's Covariance Matrix . . . 325
15.3 Partitioned Linear Model . . . 327
15.4 The Frisch–Waugh–Lovell Theorem: General Case . . . 328
15.5 The Watson Efficiency of a Subvector . . . 331
15.6 Equality of the BLUEs of $\mathbf{X}_1\boldsymbol{\beta}_1$ under Two Models . . . 333
15.7 Adding a Variable: a Recursive Formula for the BLUE . . . 335
15.8 Deleting an Observation: $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$ . . . 337
15.9 Weighted Sum of Squares of Errors . . . 338
15.10 Exercises . . . 340

16 Disjointness of Column Spaces . . . 343
16.1 Estimability of $\mathbf{X}_2\boldsymbol{\beta}_2$ . . . 345
16.2 When is $(\mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}')^{-}$ a G-inverse of $\mathbf{V}$? . . . 346
16.3 Usual Constraints . . . 347
16.4 Exercises . . . 348

17 Full Rank Decomposition . . . 349
17.1 Some Properties of an Idempotent Matrix . . . 350
17.2 Rank Additivity . . . 351
17.3 Cochran's Theorem: a Simple Version . . . 352
17.4 Proof of the RCR Using the Full Rank Decomposition . . . 354
17.5 Exercises . . . 354

18 Eigenvalue Decomposition . . . 357
18.1 The Maximal Value of $\mathbf{x}'\mathbf{A}\mathbf{x}$ . . . 360
18.2 Principal Components . . . 362
18.3 Eigenvalues of $\boldsymbol{\Sigma}_{2\times 2}$ . . . 364
18.4 Eigenvalues of the Intraclass Correlation Matrix . . . 366
18.5 Proper Eigenvalues; $\mathbf{B}$ Positive Definite . . . 367
18.6 Statistical Distance . . . 372
18.7 Proper Eigenvalues; $\mathbf{B}$ Nonnegative Definite . . . 374
18.8 Eigenvalues of the BLUE's Covariance Matrix . . . 378
18.9 Canonical Correlations and Proper Eigenvalues . . . 379
18.10 More about Canonical Correlations and the Watson Efficiency . . . 382

18.11 Exercises 386

19 Singular Value Decomposition 391
19.1 Maximal Value of $\mathbf{x}'\mathbf{A}\mathbf{y}$ 395
19.2 When is $\mathbf{A}'\mathbf{A} = \mathbf{B}'\mathbf{B}$? 396
19.3 Approximation by a Matrix of Lower Rank 397
19.4 Principal Component Analysis 402
19.5 Generalized Inverses through the SVD 407
19.6 Canonical Correlations and the SVD 408
19.7 Exercises 409

20 The Cauchy–Schwarz Inequality 415
20.1 Specific Versions of the Cauchy–Schwarz Inequality 416
20.2 Maximizing $\text{cor}(y, \mathbf{b}'\mathbf{x})$ 417
20.3 The Kantorovich Inequality 418
20.4 The Wielandt Inequality 419
20.5 How Deviant Can You Be? 420
20.6 Two Inequalities Related to Cronbach's *alpha* 422
20.7 Matrix Versions of the Cauchy–Schwarz and Kantorovich Inequality 423
20.8 Exercises 425

Notation 427

List of Figures and Philatelic Items 435

References 439

Author Index 469

Subject Index 475