INDEX

Page numbers followed by f and t indicate figures and table respectively.

Active set, defined, 36
Adaptive Lasso, 38–39
Adaptive sparse principal component analysis (ASPCA), 117
algorithm, 117
Additive penalty, 109
Akaike information criterion (AIC), 17
Alternating least-squares (ALS), 12
Ante-dependence Models, 123. See also Covariance
Approximate factor model, defined, 149
Asymptotic covariance matrix, 84
Autoregressive (AR) models, 53
Banding, defined, 141
Bartlett’s decomposition, 67
Bayes estimators, 28, 104
Bayesian analysis, of covariance matrices, 91–94
Bayesian covariance estimation, 93
Bayesian information criterion (BIC), 17, 134
Bias–variance tradeoff, 100
Bingham distributions, 92
Blood oxygenation level dependent (BOLD), 24
Canonical correlation analysis (CCA), 154, 170
Capital asset pricing model (CAPM), 68
Cauchy–Schwarz inequality, 107
Cholesky decomposition, 3, 161
GLM via, 76–79
modified, 63–67, 67t
and regression, 13–15
Cholesky factor, 137
penalizing, 132–136 (see also Sparse Gaussian graphical models)
Coherence, 30
Common factors, 15
Communality, defined, 68
Compound symmetry (CS), 53
Compressed sensing, 41–42, 75
Condition number (CN), 52, 104

© 2013 John Wiley & Sons, Inc. Published 2013 by John Wiley & Sons, Inc.

181
Constrained $L_1$-minimization for inverse matrix estimation (CLIME), 139
Conventional multivariate analysis techniques, 23
Coordinate descent algorithm, 9
Correlation matrices, priors on, 93
Covariance matrices
   Bayesian analysis of, 91–94
   characteristics of, 45
   definition and properties of, 45–49
   GLM for, 73–76
   structured, 53–56
matrix, 12, 99
   Cholesky factor of, 45
   defined, 46
   functions of, 56–60
   regression model, 75
   regularizing, 30–32
   selection models, 74, 122–124 (see also Sparse Gaussian graphical models)
Cumulative percentage of explained variance (CPEV), 111
Curds and Whey (C&W) method, 155
Data matrix, 22–26, 102
   defined, 22
Data sparsity, 21, 103
Davis and Kahan Sinθ Theorem, 53
Eckart–Young Theorem, 108, 169
Eigenvalue–eigenvector decomposition, 49
   pair, 62
Eigenvalues, shrinking the, 100–105
Eigenvectors, 99
   regularizing the, 105–107
Empirical spectral distribution, 29
Entropy loss, 27
Euclidean norm, 41
Factor
   defined, 154
   estimation and selection, 158
   scores, 72–73
   stochastic volatility model, 17
Frobenius norm, 9, 27, 100, 137
Fruit fly mortality (FFM) rate, 84–89, 85t, 86f, 87f, 88t
Functional Data, 22
   Functional magnetic resonance imaging (fMRI), 3, 24
Fused graphical Lasso (FGL), 139
Gaussian graphical model, 3, 13, 99, 121
Gaussian sequence model, 31, 117
Generalized estimating equations (GEE), 78
Generalized inverse Wishart (GIW), 91, 93
Generalized linear models (GLM), 45, 74
   for covariance matrices, 73–76
   for incomplete longitudinal data
   incoherency problem, 79–81
   incomplete data and EM algorithm, 81–84
   via Cholesky decomposition, 76–79
Gram matrix, 47
Graphical Lasso algorithm, 128–129
Group graphical Lasso (GGL), 139
Gumbel distribution, 30
Hard-thresholding function, 8
Heywood case, 71
High-dimensional data, 3–5, 21, 23, 37
High-dimensionality, 4, 30, 136
Huynh–Feldt Structure, 54
Hyperspectral Imagery, 23
Idiosyncratic errors, 15
Innovariogram, defined, 77
Intraday electricity prices, 163–165, 164f, 165f, 165t
Inverse Wishart (IW), 91
   distribution, 93
James–Stein estimator, 31
Joint graphical Lasso, defined, 139
Joint graphical models, 137–139. See also Sparse Gaussian graphical models
Joint regularization, 160–163
Karush–Kuhn–Tucker (KKT), 35
Kronecker product, 161
Kullback–Liebler divergence, 27
Lagrange’s method, 62
LARS algorithm, 37
Lasso
   adaptive, 38–39
   degrees of freedom and BIC, 37–38
INDEX

<table>
<thead>
<tr>
<th>Method</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent factor models</td>
<td>15–17, 67–73</td>
</tr>
<tr>
<td>Least angle regression (LAR) algorithm</td>
<td>34</td>
</tr>
<tr>
<td>Least squares and regularized regression</td>
<td>4–6</td>
</tr>
<tr>
<td>Linear Covariance Models (LCM)</td>
<td>74–75</td>
</tr>
<tr>
<td>Linear factor regression</td>
<td>154, 157</td>
</tr>
<tr>
<td>Linear mixed models</td>
<td>See Latent factor models</td>
</tr>
<tr>
<td>Linear pencil, defined</td>
<td>75</td>
</tr>
<tr>
<td>Log-linear covariance models</td>
<td>75–76</td>
</tr>
<tr>
<td>Markov Chain Monte Carlo (MCMC)</td>
<td>91</td>
</tr>
<tr>
<td>Matricial logarithm, defined</td>
<td>75</td>
</tr>
<tr>
<td>Matrix of covariates</td>
<td>81</td>
</tr>
<tr>
<td>Maximum likelihood estimation (MLE)</td>
<td>30, 72, 76, 153</td>
</tr>
<tr>
<td>Meinshausen and Bühlmann method (MB)</td>
<td>130</td>
</tr>
<tr>
<td>Model sparsity</td>
<td>21</td>
</tr>
<tr>
<td>Modified Cholesky decomposition</td>
<td>63–67, 67t</td>
</tr>
<tr>
<td>Moving average (MA)</td>
<td>53</td>
</tr>
<tr>
<td>Multiplicative penalty</td>
<td>109</td>
</tr>
<tr>
<td>Multivariate data</td>
<td>22, 26</td>
</tr>
<tr>
<td>Multivariate regression</td>
<td>153–154</td>
</tr>
<tr>
<td>bridge regression</td>
<td>159</td>
</tr>
<tr>
<td>group-wise penalty functions</td>
<td>159</td>
</tr>
<tr>
<td>joint regularization of B, Ω, 160–163 and LS estimators, 154–156</td>
<td></td>
</tr>
<tr>
<td>regularized estimation of B</td>
<td>158–160</td>
</tr>
<tr>
<td>Multivariate regression with covariance estimation (MRCE)</td>
<td>154</td>
</tr>
<tr>
<td>algorithm, approximate</td>
<td>162</td>
</tr>
<tr>
<td>implementing</td>
<td>162</td>
</tr>
<tr>
<td>intraday electricity prices</td>
<td>163–165</td>
</tr>
<tr>
<td>predicting asset returns</td>
<td>165–167, 166f, 167t, 168t</td>
</tr>
<tr>
<td>sparse precision matrix estimate using</td>
<td>167t</td>
</tr>
<tr>
<td>Mutual incoherence property (MIP)</td>
<td>42</td>
</tr>
<tr>
<td>Netflix data</td>
<td>21, 24, 26</td>
</tr>
<tr>
<td>Newton–Raphson algorithm</td>
<td>78, 83</td>
</tr>
<tr>
<td>Node-wise regression</td>
<td>13</td>
</tr>
<tr>
<td>Nonnegative definite</td>
<td>45, 46</td>
</tr>
<tr>
<td>matrix, 16, 52, 56, 75</td>
<td></td>
</tr>
<tr>
<td>One-dependent covariance structure</td>
<td>54</td>
</tr>
<tr>
<td>Oracle inequality</td>
<td>37</td>
</tr>
<tr>
<td>Ordinary scatterplot matrix (OSM)</td>
<td>77</td>
</tr>
<tr>
<td>Orthogonal factor models and partial correlations</td>
<td>73</td>
</tr>
<tr>
<td>Orthogonal iterations</td>
<td>118</td>
</tr>
<tr>
<td>Orthogonal matrix</td>
<td>50</td>
</tr>
<tr>
<td>Parsimony</td>
<td>4</td>
</tr>
<tr>
<td>Partial autocorrelation matrix</td>
<td>89, 90</td>
</tr>
<tr>
<td>Partial correlation coefficient</td>
<td>126</td>
</tr>
<tr>
<td>Partial regression on intervenors scatterplot matrix (PRISM)</td>
<td>77</td>
</tr>
<tr>
<td>Partial variance</td>
<td>124–125</td>
</tr>
<tr>
<td>Penalized likelihood and graphical lasso, 126–130, 131f–133f. See also Sparse Gaussian graphical models</td>
<td></td>
</tr>
<tr>
<td>Penalized matrix decomposition (PMD)</td>
<td>109</td>
</tr>
<tr>
<td>Penalized multivariate analysis (PMA)</td>
<td>107</td>
</tr>
<tr>
<td>Penalized quasi-likelihood formulation, 131–132. See also Sparse Gaussian graphical models</td>
<td></td>
</tr>
<tr>
<td>Pitprops dataset</td>
<td>111</td>
</tr>
<tr>
<td>Positive definite</td>
<td>10, 48, 53, 57, 74, 89, 129, 136, 144</td>
</tr>
<tr>
<td>covariance matrix</td>
<td>14, 46, 63, 64</td>
</tr>
<tr>
<td>Precision matrix</td>
<td>16, 121</td>
</tr>
<tr>
<td>Principal component analysis (PCA)</td>
<td>3, 112t</td>
</tr>
<tr>
<td>consistency of</td>
<td>114–117</td>
</tr>
<tr>
<td>duality between SVD and</td>
<td>107–110</td>
</tr>
<tr>
<td>implementing sparse, 110–111</td>
<td></td>
</tr>
<tr>
<td>sparse, 10–12, 106</td>
<td></td>
</tr>
<tr>
<td>variance property</td>
<td>61–63</td>
</tr>
<tr>
<td>Principal components (PC)</td>
<td>10</td>
</tr>
<tr>
<td>Principal subspace estimation</td>
<td>118</td>
</tr>
<tr>
<td>Proto correlation matrix</td>
<td>89</td>
</tr>
<tr>
<td>Quadratic loss function</td>
<td>27</td>
</tr>
<tr>
<td>Random correlation matrices, simulating, 89–91</td>
<td></td>
</tr>
<tr>
<td>Reduced rank regression (RRR)</td>
<td>153, 156–158</td>
</tr>
<tr>
<td>Netflix data</td>
<td>21, 24, 26</td>
</tr>
<tr>
<td>Newton–Raphson algorithm</td>
<td>78, 83</td>
</tr>
</tbody>
</table>
Regression
  and Cholesky decomposition, 13–15
  coefficients and residual variances, 67t
  Regressograms, 77
  Relevance networks, 23
  Restricted isometry property (RIP), 42
  Ridge regression, 5

Sample correlation matrix, 104
Sample covariance matrix, 100, 105, 127
  banding, 142–143
  shrinking, 26–29, 27f
  tapering, 144
  thresholding, 9–10, 145–148
Sample eigenvalues, distribution of, 29–30
SAS PROC MIXED software, 75
Schur formula, 129
Schur Theorem, 47
SCoTLASS, 106
Simplified component technique (SCoT), 106
Semidefinite programming (SDP), 140
Singular value decomposition (SVD), 8, 107, 154
  duality between PCA and, 107–110
Smoothly clipped absolute deviation (SCAD), 9, 39
  penalty, 19
  properties, 39–40
Soft-thresholding operator, 7, 33
  defined, 8
  plot of, 7f
Sparse covariance matrices, low-rank plus, 149–150
Sparse Gaussian graphical models, 121–122
  Cholesky factor, penalizing the, 132–136
  consistency and sparsistency, 136–137
  covariance selection models, 122–124
  joint graphical models, 137–139
  penalized likelihood and graphical lasso, 126–130
  penalized quasi-likelihood formulation, 131–132
  regression interpretation of entries of, 124–126
Sparsely initialized thresholded subspace iterations algorithm, 118
Sparse PCA
  method, 106
  and regression, 10–12
Sparse principal component (SPC), 107, 112t
  algorithm, 107
Sparse pseudo-likelihood inverse covariance estimation (SPLICE) algorithm, 131
Sparse singular value decomposition (SSVD), 100, 112–114
  kth layer, 113
Sparsistency, 136–137
Sparsity, 30, 100
  index, 4, 36
  principle, 4
Spatial autoregressive (SAR) models, 76
Spatial Data, 22
Specific variance, defined, 68
Spectral decomposition, 49–53
Spiked covariance model, 114, 119. See also Principal component analysis (PCA)
Square matrix, 49
  \( \Sigma^{-1} \), regression interpretation of, 124–126.
  See also Sparse Gaussian graphical models
  partial correlation interpretation, 126
Star-shaped model, defined, 122. See also Covariance
Sure independence screening (SIS), 34
Symmetric matrix, 48
Tapering, defined, 141
Tickby-Tick Financial Data, 23
Time Series Data, 22
Toeplitz matrix, 94
Tracey–Widom distribution, 29
Transposable Data. See Spatial Data
Transposable matrices, 23
True negative rate (TNR), 163
True positive rate (TPR), 163
Tuning-insensitive graph estimation and regression (TIGER), 18, 130
Universal thresholding, defined, 147
Varimax criterion, 70
Vector autoregressive (VAR), 166
Web Term-Document Data, 23
Weighted least-squares (WLS), 72, 153
Weyl’s Theorem, 52