

**Editors-in-Chief:**

J.-M. Morel, Cachan

B. Teissier, Paris

**Advisory Board:**

Camillo De Lellis, Zürich

Mario di Bernardo, Bristol

Alessio Figalli, Austin

Davar Khoshnevisan, Salt Lake City

Ioannis Kontoyiannis, Athens

Gabor Lugosi, Barcelona

Mark Podolskij, Aarhus

Sylvia Serfaty, Paris and New York

Catharina Stroppel, Bonn

Anna Wienhard, Heidelberg

# Saint-Flour Probability Summer School

The Saint-Flour volumes are reflections of the courses given at the Saint-Flour Probability Summer School. Founded in 1971, this school is organised every year by the Laboratoire de Mathématiques (CNRS and Université Blaise Pascal, Clermont-Ferrand, France). It is intended for PhD students, teachers and researchers who are interested in probability theory, statistics, and in their applications.



The duration of each school is 13 days (it was 17 days up to 2005), and up to 70 participants can attend it. The aim is to provide, in three high-level courses, a comprehensive study of some fields in probability theory or Statistics. The lecturers are chosen by an international scientific board. The participants themselves also have the opportunity to give short lectures about their research work.

Participants are lodged and work in the same building, a former seminary built in the 18th century in the city of Saint-Flour, at an altitude of 900 m. The pleasant surroundings facilitate scientific discussion and exchange.

The Saint-Flour Probability Summer School is supported by:

- Université Blaise Pascal
- Centre National de la Recherche Scientifique (C.N.R.S.)
- Ministère délégué à l'Enseignement supérieur et à la Recherche

For more information, see

**<http://recherche.math.univ-bpclermont.fr/stflour/stflour-en.php>**

**Christophe Bahadoran**

bahadora@math.univ-bpclermont.fr

**Arnaud Guillin**

Arnaud.Guillin@math.univ-bpclermont.fr

**Laurent Serlet**

Laurent.Serlet@math.univ-bpclermont.fr

Université Blaise Pascal – Aubière cedex, France

More information about this series at <http://www.springer.com/series/304>

Sara van de Geer

# Estimation and Testing Under Sparsity

École d'Été de Probabilités de Saint-Flour  
XLV – 2015

Sara van de Geer  
Seminar für Statistik HGG 24.1  
ETH Zentrum  
Zürich, Switzerland

ISSN 0075-8434                      ISSN 1617-9692 (electronic)  
Lecture Notes in Mathematics  
ISBN 978-3-319-32773-0              ISBN 978-3-319-32774-7 (eBook)  
DOI 10.1007/978-3-319-32774-7

Library of Congress Control Number: 2016943427

Mathematics Subject Classification (2010): 62-XX; 60-XX, 68Q87

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG Switzerland

*To my mother*



# Preface

These lecture notes were written for the 45th École d'Été de Probabilités de Saint-Flour in 2015. It was a great honour and a pleasure for me to be invited there and enjoy the *Grand Séminaire*, the beautiful surroundings and, most of all, the wonderful participants. The summer was very warm, with temperatures rising up to 40°. However, the old building kept us cool and ready to digest quite a few lectures each day. Indeed, the amount of mathematical activity was impressive, with a scientific programme in both mornings and afternoons and many excellent talks presented by my fellow participants. But there were also numerous other activities: chatting, hiking, dining, ping-pong, and looking after the small kids.

The notes aim to provide an overview of the techniques used to obtain theoretical results for models of high-dimensional data. High-dimensional models have more unknown parameters  $p$  than observations  $n$ . Sparsity means that the number of relevant—or active—parameters is actually much smaller than  $n$ . However, which parameters these are, or how many there are, is not known beforehand. The first goal in these notes is to study methods that perform almost as well as in the hypothetical case with a known active set. The next goal is then to zoom in on certain parameters of interest and test their significance.

An important technique in high-dimensional regression is the Lasso method. It is taken here as a prototype for understanding other methods, such as those inducing structured sparsity or low rank or those based on more general loss functions. The common features are highlighted so that—hopefully—they will serve as a good starting point for the theory of new methods, not treated in this book.

I am very grateful to the Scientific Board for having given me the opportunity to lecture at the Saint-Flour Summer School. I thank all participants and am greatly indebted to Claire Boyer, Yohann De Castro and Joseph Salmon, who spontaneously did a careful reading of the version of the lecture notes available at the time. The mistakes in the current version have entered *after* their proofreading. Special thanks goes to the organisers on the spot, Christophe Bahadoran and Laurent Serlet. I also thank the two other main lecturers, Sourav Chatterjee and Lorenzo Zambotti for our

very pleasant time together and for their inspiring courses which opened up new windows with magnificent views.

I thank my colleagues at the Seminar for Statistics in Zürich for their support, for their motivating interest and for providing me with an ideal research environment.

Finally, a very special thanks is extended to my family.

Zürich, Switzerland  
March 2016

Sara van de Geer



# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Regularized Empirical Risk Minimization	1
1.2	Organization of the Book	3
1.3	How to Read This Book	4
<b>2</b>	<b>The Lasso</b>	5
2.1	The Linear Model with $p < n$	5
2.2	The Linear Model with $p \geq n$	7
2.3	Notation	8
2.4	The Lasso, KKT and Two Point Inequality	8
2.5	Dual Norm and Decomposability	10
2.6	Compatibility	11
2.7	A Sharp Oracle Inequality	12
2.8	Including a Bound for the $\ell_1$ -Error and Allowing Many Small Values	13
2.9	The $\ell_1$ -Restricted Oracle	17
2.10	Weak Sparsity	18
2.11	Complements	20
	2.11.1 A Bound Based on an Alternative Method of Proof	20
	2.11.2 When There Are Coefficients Left Unpenalized	21
	2.11.3 A Direct Proof of Theorem 2.1	21
	Problems	23
<b>3</b>	<b>The Square-Root Lasso</b>	27
3.1	Introduction	27
3.2	KKT and Two Point Inequality for the Square-Root Lasso	28
3.3	A Proposition Assuming No Overfitting	28
3.4	Showing the Square-Root Lasso Does Not Overfit	29
3.5	A Sharp Oracle Inequality for the Square-Root Lasso	32
3.6	A Bound for the Mean $\ell_1$ -Error	33

3.7	Comparison with Scaled Lasso .....	35
3.8	The Multivariate Square-Root Lasso .....	36
	Problems .....	38
<b>4</b>	<b>The Bias of the Lasso and Worst Possible Sub-directions .....</b>	<b>41</b>
4.1	Introduction .....	41
4.2	Matrix Algebra .....	42
4.3	Notation .....	44
4.4	Bias and Sup-Norm Bounds for the Lasso with Random Design .....	44
4.5	Bias and Sup-Norm Bounds for the Lasso with Fixed Design .....	47
4.6	Oracle Inequalities for Groups .....	49
4.7	Worst Possible Sub-directions and the Irrepresentable Condition .....	53
4.8	De-selection of Variables in the Approximately Sparse Case .....	54
4.9	A Semi-parametric Lower Bound .....	56
	Problems .....	59
<b>5</b>	<b>Confidence Intervals Using the Lasso .....</b>	<b>61</b>
5.1	A Surrogate Inverse for the Gram Matrix .....	61
5.2	Asymptotic Linearity of the De-sparsified Lasso .....	63
5.3	Confidence Intervals for Linear Contrasts .....	67
5.4	Confidence Sets for Groups .....	68
5.5	Random Design with Known Co-variance Matrix .....	70
	Problems .....	73
<b>6</b>	<b>Structured Sparsity .....</b>	<b>75</b>
6.1	The $\Omega$ -Structured Sparsity Estimator .....	75
6.2	Dual Norms and KKT-Conditions for Structured Sparsity .....	76
6.3	Two Point Inequality .....	77
6.4	Weak Decomposability and $\Omega$ -Triangle Property .....	78
6.5	$\Omega$ -Compatibility .....	80
6.6	A Sharp Oracle Inequality with Structured Sparsity .....	81
6.7	Norms Stronger than $\ell_1$ .....	83
6.8	Structured Sparsity and Square-Root Loss .....	83
6.8.1	Assuming There Is No Overfitting .....	84
6.8.2	Showing There Is No Overfitting .....	85
6.8.3	A Sharp Oracle Inequality .....	85
6.9	Norms Generated from Cones .....	86
6.10	$\Omega_*$ -Bounds for the $\Omega$ -Structured Sparsity Estimator .....	90
6.11	Confidence Intervals Using Structured Sparsity .....	92
6.12	Complements .....	96
6.12.1	The Case Where Some Coefficients Are Not Penalized .....	96
6.12.2	The Sorted $\ell_1$ -Norm .....	96
6.12.3	A Direct Proof of Theorem 6.1 .....	98
	Problems .....	99

<b>7</b>	<b>General Loss with Norm-Penalty</b>	103
7.1	Introduction	103
7.2	Two Point Inequality, Convex Conjugate and Two Point Margin	104
7.3	Triangle Property and Effective Sparsity	106
7.4	Two Versions of Weak Decomposability	108
7.5	A Sharp Oracle Inequality	110
7.6	Localizing (or a Non-sharp Oracle Inequality)	113
	Problems	117
<b>8</b>	<b>Empirical Process Theory for Dual Norms</b>	121
8.1	Introduction	121
8.2	The Dual Norm of $\ell_1$ and the Scaled Version	121
8.3	Dual Norms Generated from Cones	124
8.4	A Generalized Bernstein Inequality	124
8.5	Bounds for Weighted Sums of Squared Gaussians	126
8.6	The Special Case of Chi-Squared Random Variables	128
8.7	The Wedge Dual Norm	129
	Problems	130
<b>9</b>	<b>Probability Inequalities for Matrices</b>	133
9.1	Hoeffding and Bernstein	133
9.2	Matrix Hoeffding and Matrix Bernstein	134
9.3	Matrix Inequalities for Gaussian and Rademacher Sums	135
9.4	Matrices with a Finite Orlicz Norm	135
9.5	An Example from Matrix Completion	136
	Problems	137
<b>10</b>	<b>Inequalities for the Centred Empirical Risk and Its Derivative</b>	139
10.1	Introduction	139
10.2	Exponential Families	140
10.2.1	Density Estimation	140
10.2.2	Regression with Fixed Design	141
10.3	Projection Estimators	142
10.4	The Linear Model with Random Design	143
10.4.1	Known Co-variance Matrix	143
10.4.2	Unknown Co-variance Matrix	144
10.5	Generalized Linear Models	146
10.6	Other Models	148
	Problems	149
<b>11</b>	<b>The Margin Condition</b>	151
11.1	Introduction	151
11.2	Two Term Taylor Approximations	153
11.3	Exponential Families	155
11.3.1	Density Estimation	155
11.3.2	Regression with Fixed or Random Design	157
11.3.3	Least Squares Regression with Fixed Design	158

11.4	Projection Estimators .....	158
11.5	The Linear Model with Random Design .....	159
11.5.1	Known Co-variance Matrix .....	159
11.5.2	Unknown Co-variance Matrix .....	160
11.6	Generalized Linear Models .....	160
11.7	The Two Point Margin Condition and Effective Sparsity .....	161
11.8	Other Models .....	164
	Problems .....	164
<b>12</b>	<b>Some Worked-Out Examples</b> .....	<b>167</b>
12.1	The Lasso and Square-Root Lasso Completed .....	167
12.2	Least Squares Loss with $\Omega$ -Structured Sparsity Completed .....	169
12.3	Density Estimation Using Projections .....	172
12.4	Logistic Regression .....	174
12.4.1	Logistic Regression with Fixed, Bounded Design .....	175
12.4.2	Logistic Regression with Random, Unbounded Design .....	176
12.5	Trace Regression with Nuclear Norm Penalization .....	182
12.5.1	Some Useful Matrix Inequalities .....	182
12.5.2	Dual Norm of the Nuclear Norm and Its Triangle Property .....	183
12.5.3	An Oracle Result for Trace Regression with Least Squares Loss .....	186
12.5.4	Robust Matrix Completion .....	187
12.6	Sparse Principal Components .....	190
12.6.1	Two Point Margin and Two Point Inequality for Sparse PCA .....	192
12.6.2	Effective Sparsity and Dual Norm Inequality for Sparse PCA .....	194
12.6.3	A Sharp Oracle Inequality for Sparse PCA .....	195
	Problems .....	196
<b>13</b>	<b>Brouwer's Fixed Point Theorem and Sparsity</b> .....	<b>199</b>
13.1	Brouwer's Fixed Point and Schauder's Extension .....	199
13.2	$\Omega_*$ -Bound for $\Omega$ -Structured Sparsity M-Estimators .....	200
13.3	The Case of No Penalty .....	202
13.4	The Linear Model with Random Design .....	203
13.5	De-selection of Variables and the Irrepresentable Condition .....	204
13.6	Brouwer and Exponential Families .....	209
13.7	The Graphical Lasso .....	210
	Problems .....	214
<b>14</b>	<b>Asymptotically Linear Estimators of the Precision Matrix</b> .....	<b>215</b>
14.1	Introduction .....	215
14.2	The Case $p$ Small .....	216

14.3	The Graphical Lasso .....	216
14.4	The Square-Root Node-Wise Lasso .....	218
	Problems .....	220
<b>15</b>	<b>Lower Bounds for Sparse Quadratic Forms</b> .....	<b>223</b>
15.1	Introduction .....	223
15.2	Lower Bound .....	225
15.3	Application to Effective Sparsity .....	229
	Problems .....	230
<b>16</b>	<b>Symmetrization, Contraction and Concentration</b> .....	<b>233</b>
16.1	Symmetrization and Contraction .....	233
16.2	Concentration Inequalities .....	235
16.2.1	Massart's Inequality .....	236
16.2.2	Bousquet's Inequality .....	236
	Problems .....	237
<b>17</b>	<b>Chaining Including Concentration</b> .....	<b>239</b>
17.1	Introduction .....	239
17.2	Notation .....	240
17.3	Hoeffding's Inequality .....	241
17.4	The Maximum of $p$ Random Variables .....	243
17.5	Expectations of Positive Parts .....	244
17.6	Chaining Using Covering Sets .....	245
17.7	Generic Chaining .....	246
17.8	Bounding the Expectation .....	247
17.9	Concentration .....	250
	Problems .....	252
<b>18</b>	<b>Metric Structure of Convex Hulls</b> .....	<b>255</b>
18.1	Introduction .....	255
18.2	Possible Generations .....	257
18.3	A Conjecture .....	258
18.4	Dependence on the Maximal Eigenvalue .....	259
18.5	Approximation Using Maurey's Lemma .....	260
18.6	Dependence on the Entropy .....	261
18.7	Some Further Bounds .....	263
	Problems .....	265
	<b>References</b> .....	<b>267</b>
	<b>Author Index</b> .....	<b>271</b>
	<b>Index</b> .....	<b>273</b>