Systematic Analysis in Three-Dimensional Reconstruction of Weak-Lensing Mass Maps with a Sparsity Prior

Shouzhuo Yang

Advisor: Naoki Yoshida

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Engineering

Department of Operations Research and Financial Engineering
Princeton University

May 2023

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Shouzhuo Yang

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Shouzhuo Yang

Abstract

1. cosmology basics... ΛCDM. A little bit of FDR metric...

2. Dark Matter halos. virialization... growth of perturbations. Different kinds of dark matter halo. Describe very briefly different dark matter prediction of dark matter halo profiles.

3. Weak lensing basics. Some Newtonian/GR level calculation. What can the weak lensing probe about our universe.

4. Weak lensing systematics. This is the part I am worst at.. I will need to ask Yoshida about whether to write about it. Weak lensing noise is something I definitely should mention.

5. Sparsity analysis

6. 3d mass reconstruction (including a mention of other reconstruction research).

7. results?

Acknowledgements

I wish to first give my heartfelt gratitude toward the Department of the Physics and Astronomy for hosting the honors program and provide me this opportunity to summarize my work in weak lensing cosmology. It has been truly a satisfying process. My thanks also goes to Professor Tristan Smith for setting up this project with Professor Naoki Yoshida, who offered me a wonderful journel of cosmology. Also, I wish mention my collaborator and friend Xiangchong Li from Carnegie Mellon University who helped me not only with this project but various aspect on physics. I appreciate the help also of my two other research advisors, Peter Collings and Michael Brown, for supporting my thesis in their own way. Finally, I wish to thank my partner Meihan "Della" Guo, who, though not contributing scientifically, provided me emotional support that have gotten me through the project.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Hello World. See [3].

See [3] and [2] for more information.

This is the introduction of my sample thesis. Let me just state that this is a very, very limited introduction to LaTeX, and I can not do justice to it through the next few pages.

I recommend reading this .pdf file along with the .tex file open on the side so that you can compare the pure ASCII text and the final result after compilation.

1.1 A Different Word Processing System

I highly recommend the booklet "The Not So Short Introduction to LaTeX 2_ε" which is free and very well-written. It is available on:

`http://tobi.oetiker.ch/lshort/lshort.pdf`.

I refer you to the introduction there.

Chapter 2

Development of the Department of Operations Research and Financial Engineering

This Chapter has purposefully a very long title to illustrate how $\LaTeX\,2_\varepsilon$ handles such long names. Now it is a good time to look on the table of contents and see how this Chapter and also Chapter 1 are listed. Notice that the number 1 in the phrase "and also Chapter 1" is automatically generated by using the command `\ref{ch:intro}`, because we labeled Chapter 1 '`ch:intro`' by using the command `\label{ch:intro}` after the `\chapter{Introduction}` declaration.

2.1 Initial Setup

A few things here to start. And here as well, since we need to fill in at least a line. Almost. . . , there we go!

Notice that I entered ellipses after the word 'Almost' above with the command `\ldots`, and not by simply typing three dots. Compare the result here: . . . vs. ...[1]

[1] Word does that too, but lousily!

Maybe some more words in a new paragraph. And more, and more, and more. Furthermore, additionally, in addition, and so on. Notice here that the word 'Furthermore' was broken into 'Fur-' and 'thermore' in order to fit the line—Word, as stupid as it is, would simply place the entire word on the next line, thus increasing the distance between words on the first line to fill up the entire line.

2.1.1 Additional Structure: The Use of Subsections

We are in a subsection now (two levels down from a Chapter). When we refer to $X.Y.Z$, we mean Chapter X, Section Y, and subsection Z. We declare a Chapter by the command `\chapter{`*title of Chapter*`}`, a Section by `\section{`*title of Section*`}`, and so on.

The nice thing about LaTeX is that it takes care of the chapter, section, and subsection numbering automatically. If I were to add another subsection before this one the subsection number would change (increment by one). This section is 2.1.1 and I referred to it using the command `\ref{`*label of this section*`}`. I inserted a label right after the `\subsection` declaration by typing `\label{`*label of this section*`}`.

A subsubsection

Just for fun! Notice that no number is alloted for such a low level environment but it sometimes useful.

2.1.2 Another Subsection

And so on. . . .

2.2 Mathematical Symbols

Let $X = \{X_n, n \in \mathbb{N}\}$ be a Markov chain with state space $\mathcal{D}$. Throughout this thesis, we use the notation

$$p_{ij} := \mathbb{P}\{X_{n+1} = j \mid X_n = i\}, \quad i, j \in \mathcal{D} \tag{2.1}$$

for the transition probabilities of the Markov chain X. Furthermore, we denote by P the transition matrix, $P = [p_{ij}]_{i,j\in\mathcal{D}}$.

When we wrote (2.1) we implicitly assumed that the Markov chain X is time-homogeneous.

Let us also define Y,

$$Y = (Y_n)_{n=0,1,2,\ldots}$$

to be another process. Notice that the second equation does not take a number on the right—this is the use of `\begin{equation*}` environment.

Notice that the all the math characters, X, $\mathcal{D}$, and others such as α, β, γ are part of the text in LaTeX. On the contrary, Word includes such characters as foreign objects (usually images), which increases the size of the document file, sometimes makes them disappear, but most importantly are not as aesthetically pleasing as the resulting characters here.

2.3 Citing and Bibliography

When working with large documents you need an easy way to cite your references without having to go back to your list all the time to remember the names of the authors and the year of publication. Even more importantly, you need to have all your references listed in the end of the document in alphabetical order. Of course, they all need to be syntactically the same so that alone makes the manual entry of references a big pain. Thankfully, LaTeX takes care of that in a very easy and elegant way, using BibTeX.

I cite here a few books, papers, and technical reports, and please go to page 16 to see the resulting bibliography.

According to the books by [8], [5], and [11] and the articles by [10], [7], and [9] we conclude absolutely nothing. However, in his report, [1] claims that otherwise. All these citations were entered by `\cite{`*citation label*`}`.

Notice the different citation style that follows: it is parenthetical, and observe that only one pair of parentheses is required [4, see Theorem 5.2 on pg. 32]. This citation is entered by typing `\cite[see Theorem 5.2 on pg. 32]{AMM05}` in the `.tex` file. (Here, the citation label corresponding to [4] is obviously `AMM05`.)

The citations are included in the file `refs.bib` under the folder `Bibliography`. You can modify it and make your own references.

Also notice that LaTeX, by default, includes in the Bibliography section only the references you actually cited throughout the text. If you want a source to appear in the Bibliography section without actually citing it anywhere in your text use the command `\nocite{`*citation label*`}`. For example here I type `\nocite{B95}` and you see no citation appear—however look at the fourth entry of the Bibliography. That cited book does not appear anywhere in this thesis, other than the Bibliography.

2.4 Referencing Figures and Tables

The very informative Figure 3.1 is on page 11. Both of these numbers were automatically generated—which is great when you add a new figure before the one you just inserted, because the numbering changes automatically for you. Use `\ref{fig:dens}` for the figure number and `\pageref{fig:dens}` for the page number where the figure is located. Here, `fig:dens` was the label of the figure (see actual `.tex` file for more information). Remember that LaTeX does not work like Word—the figures and tables are **not** always placed exactly where you want them, so avoid writing "according to the figure below...," and prefer writing "according to Figure [*figure number*]...," instead. The same things go unchanged for tables. Notice that when I talk about figures and tables in general, I do not need to capitalize them, however if I talk specifically about Figure 3.1 and Table 3.1, I'd better respect them and capitalize the 'f' and the 't.'

Since we're at it, notice that the quotes ', ', ", " are not inserted like in Word. For ' you need to use the ` key that is located above the `Tab` button. For ' you just press the ' key, exactly to the left of the `Enter` key. For double quotes just double the appropriate single quotes without leaving any space.

2.5 Introduction to Sparisty Analysis

Given the prevalence and the strength of noise in weak lensing analysis, to successfully reconstruct physical information from weak lensing data, one may be prompted to use *sparsity*. A *sparse* model indicates only a relatively small amount of parameters controls the behavior of the system. In weak lensing and specifically 3D reconstruction, we apply this idea to assert that galaxy clusters are only sparsely distributed in the present universe.

To implement the sparisty prior in an analysis, first consider the familiar linear regression model:

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + e_i \tag{2.2}$$

where β_0 and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ are unknown multidimensional linear parameters of some model. y_i's are measured values, and e_i's are errors of estimation.

To retrieve the best predicted model, one acquire an estimation of β parameters by mininizing the least-squares objective function.

$$\text{minimize}_{\beta,\beta_0} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{2.3}$$

However, when the number of undertermined parameters is large, such an objective function may yield nonzero value for the β vector and yield degenerate solutions. This makes the interpretation of the result difficult. A solution is just to *regularize* the estimation process. The regime we will be working with in this work is called the Lasso regularization, which uses the Lasso Estimator:

$$\text{minimize}_{\beta_0,\beta} \{ \frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \} \tag{2.4}$$

subject to

$$\|\beta\|_1 \leq t.$$

Where $\|\beta\|_1$ represents the l_1-norm which is the sum of the absolute values of each entries in β.

In matrix form, one may have

$$\text{minimize}_{\beta_0,\beta} \left\{ \frac{1}{2N} \|\mathbf{y} - \beta_{\mathbf{0}}\mathbf{1} - \mathbf{X}\beta\|_{\mathbf{2}}^{\mathbf{2}} \right\} \tag{2.5}$$

subject to

$$\|\beta\|_1 \leq t.$$

One can also rewrite this into the Lagrangian form

$$\text{minimize}_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|_{\mathbf{2}}^{\mathbf{2}} + \lambda \|\beta\|_{\mathbf{1}} \right\} \tag{2.6}$$

Here then $\lambda \geq 0$ now controls the degree of sparisty we impose on the minimization, with larger λ correspond to stronger sparse condition. For our application, we will use a slightly modified version of the Lasso regression regime where λ is normalized with respect to standard deviation of the underlying shear field estimation.

2.6 Adaptative Lasso Regression

In this section we describe the Adaptive Lasso Regression algorithm introduced in [12]. In (Fan and LI 2001), it was discovered the lasso algorithm has biased estimates for large coefficients. Meinshausen and Bühlmann (2004) also sproved even the optimal λ value yields in consistent variable selection results, meaning that the regression algorithm sometimes picks up signal from the noise. These are all signs that the "traditional" lasso regression in Eqn. 2.6 does not satisfy the oracle property, which is the tendancy that a penalized (e.g. the Lasso l_1 norm penalty or the Smoothly Clipped Absolute Deviation penalty) regression estimator being able to recover the true underlying model consistently. The nature of the adaptive Lasso algorithm is to use an adaptive weight to penalize different projection coeffcients in the l_1 penalty.

The adaptive lasso regression algorithm can be described as follows. After getting the Lasso estimator $\hat{\beta}_{\text{Lasso}}$ by minimizing Eqn. 2.6, we then define the adaptive weight, with $\tau = 2$ in this work (although it can be any postive real number):

$$\hat{w} = \frac{1}{|\hat{\beta}_{\text{Lasso}}|^{\tau}}. \tag{2.7}$$

This adaptive weight enhance the shrinkage in for the coefficients with smaller amplitudes, as can be seen the inverse of the norm of the preliminary estimates. The adaptive Lasso estimator is then acquire by minimizing

$$\hat{\beta}' = \text{arg minimize}_{\beta' \in \mathbb{R}^p} \Big\{ \frac{1}{2} \sum \|\mathbf{y} - \mathbf{X}\beta'\|_{\mathbf{2}}^{\mathbf{2}} + \lambda_{\text{ada}} \|\hat{\mathbf{w}} \circ \beta'\|_{\mathbf{1}} \Big\}, \tag{2.8}$$

where λ_{ada} is the adaptive Lasso penalty parameter, which we have taken to be the same as λ in Eqn. 2.6. "$\circ$" refers to the element-wise product.

Chapter 3

Analysis of Problem

Time for some analysis, and probably graphs and tables. Thankfully LaTeX (and LaTeX 2$_\varepsilon$) provides very nice environments for both.

3.1 Preliminary Analysis

Assume that the random variable D given $\tilde{p} = p$ is binomially distributed with parameters 50 and p (probability of success, in this case default). We also take the cumulative distribution function of $\tilde{p}$ to be

$$F(\theta) = \mathbb{P}\{\tilde{p} \leq \theta\} = \Phi\left(\frac{1}{\rho}\left(\sqrt{1-\rho^2}\,\Phi^{-1}(\theta) - \Phi^{-1}(\bar{p})\right)\right)$$

where Φ is the cumulative standard normal distribution function, ρ is the correlation coefficient between the idiosyncratic and market factors and $\bar{p}$ is the mean default probability ($\bar{p} = \mathbb{E}\tilde{p}$). To calculate the density of $\tilde{p}$ let

$$\begin{aligned} h(\theta, \rho, \bar{p}) &:= \frac{1}{\rho}\left(\sqrt{1-\rho^2}\,\Phi^{-1}(\theta) - \Phi^{-1}(\bar{p})\right), \\ \varphi(\theta) &:= \frac{d}{d\theta}\Phi(\theta) = \frac{1}{\sqrt{2\pi}}e^{-\theta^2/2}, \end{aligned}$$

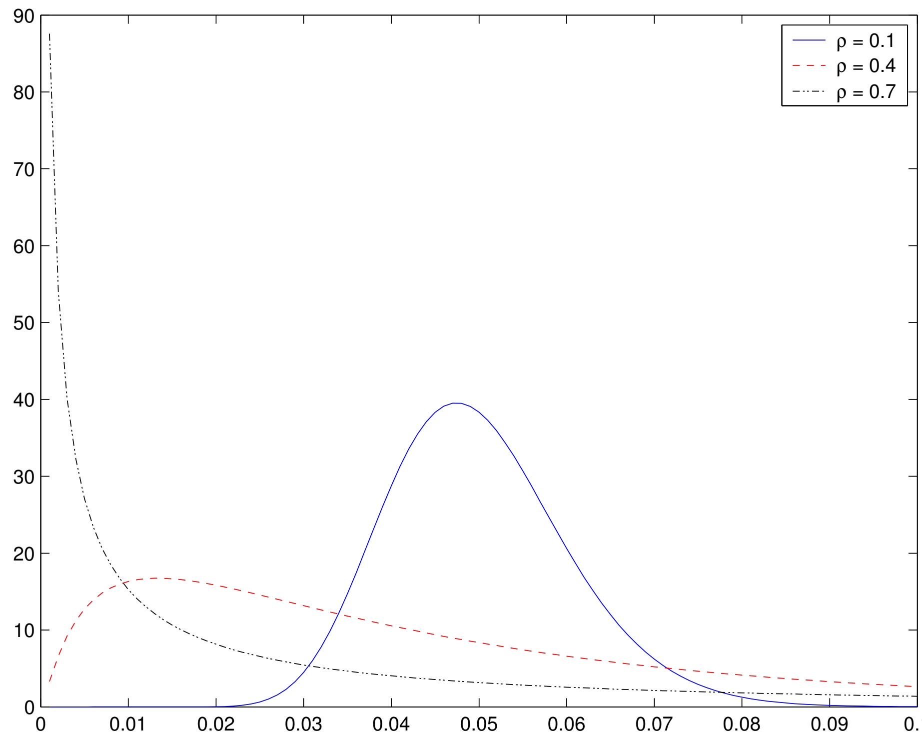

Figure 3.1: The density function f as given in (3.1) for three different ρ's and $\bar{p} = 0.05$. (Plotted on $[0, 0.1]$ for convenience.)

and notice that since Φ is a bijection we have

$$\Phi \circ \Phi^{-1}(\theta) = \Phi^{-1} \circ \Phi(\theta) = \theta,$$

for every $\theta \in \mathbb{R}$. Then, we have for the density of $\tilde{p}$,

$$\begin{aligned} f(\theta, \rho, \bar{p}) &= \frac{d}{d\theta} F(\theta) = \Phi'(h(\theta, \rho, \bar{p})) \frac{\partial}{\partial \theta} h(\theta, \rho, \bar{p}) \\ &= \varphi(h(\theta, \rho, \bar{p})) \frac{\sqrt{1-\rho^2}}{\rho} \frac{d}{d\theta} \Phi^{-1}(\theta) \\ &= \varphi(h(\theta, \rho, \bar{p})) \frac{\sqrt{1-\rho^2}}{\rho} \frac{1}{\varphi(\Phi^{-1}(\theta))}, \end{aligned} \tag{3.1}$$

for $\theta \in (0,1)$ and zero otherwise, since

$$\begin{aligned} \frac{d}{d\theta}\left(\Phi(\Phi^{-1}(\theta))\right) &= \Phi'(\Phi^{-1}(\theta))\frac{d}{d\theta}\Phi^{-1}(\theta) \Leftrightarrow \\ \frac{d}{d\theta}\theta &= \varphi(\Phi^{-1}(\theta))\frac{d}{d\theta}\Phi^{-1}(\theta) \Leftrightarrow \\ \frac{d}{d\theta}\Phi^{-1}(\theta) &= \frac{1}{\varphi(\Phi^{-1}(\theta))}. \end{aligned}$$

The density of $\tilde{p}$ is shown in Figure 3.1 for three different values of ρ and $\bar{p} = 0.05$. The effect of the correlation is to put more mass towards higher default probabilities as the correlation increases, thus resulting in larger number of defaults as the correlation increases.

Table 3.1: Values of the CDO Tranches.

		ρ		
		0.1	0.4	0.7
Equity	$C_E^\star(T)$	2.5712	2.8957	3.5642
Junior	$C_J^\star(T)$	9.9289	9.6137	9.2120
Senior	$C_S^\star(T)$	35.0000	34.9905	34.7239
Sum	$C_E^\star(T) + C_J^\star(T) + C_S^\star(T)$	47.5000	47.5000	47.5001

Using the default values for the parameters as above we get the values for the tranches in Table 3.1.

Let's see one more result on Table 3.1. According to...

Appendix A

Code

```
% Question 1
warning('off'); % To protect our nerves
p = 0.05;
K_E = 5;
K_J = 10;
K_S = 35;

tmp1 = [];
for i = 0:(K_E-1)
     tmp1(i+1) = (K_E - i) * binopdf(i,50,p);
end
C_E = sum(tmp1)

tmp2 = [];
for i = (K_E+1):50
    tmp2(i-K_E) = (i-K_E) * binopdf(i,50,p);
end
```

```
tmp3 = [];
for i = (K_E + K_J +1):50
    tmp3(i- K_E - K_J) = (i - K_E - K_J) * binopdf(i,50,p);
end
C_J = K_J - sum(tmp2) +sum(tmp3)


C_S = K_S - sum(tmp3)


% Sanity check
C_E + C_J + C_S
```

Bibliography

[1] K. Aas. Modelling the dependence structure of financial assets: A survey of four copulas. Technical report, Norwegian Computing Center, December 2004. SAMBA/22/04.

[2] E. Altman, A. Resti, and A. Sironi. Default recovery rates in credit risk modeling: A review of the literature and empirical evidence. Working paper, New York University, December 2003.

[3] M. Ammann. *Credit Risk Valuation*. Springer, 2nd edition, 2001.

[4] A. Antonov, S. Mechkov, and T. Misirpashaev. Analytical techniques for synthetic cdos and credit default risk measures. Work in progress, NumeriX, May 2005.

[5] T. R. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer, 2002.

[6] P. Billingsley. *Probability and Measure*. Wiley, 3rd edition, 1995.

[7] R. Blanco, S. Brennan, and I. W. Marsh. An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *Journal of Finance*, 60(5):2255–2281, 2005.

[8] E. Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

[9] P. Cotton, J.-P. Fouque, G. Papanicolaou, and R. Sircar. Stochastic volatility corrections for interest rate derivatives. *Mathematical Finance*, 14(2):173–200, 2004.

[10] D. Duffie and N. Gârleanu. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1):41–59, 2001.

[11] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling*. Springer, 1997.

[12] H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, 2006.