A unified view of high-dimensional bridge regression

Haolei Weng

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

A unified view of high-dimensional bridge regression

Haolei Weng

In many application areas ranging from bioinformatics to imaging, we are interested in recovering a sparse coefficient $\beta \in \mathbb{R}^p$ in the high-dimensional linear model $y = X\beta + w$, when the sample size n is comparable to or less than the dimension p. One of the most popular classes of estimators is the ℓ_q-regularized least squares (LQLS), a.k.a. bridge regression (Frank and Friedman, 1993; Fu, 1998), given by the following optimization problem:

$$\hat{\beta}(\lambda, q) \in \arg\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \sum_{i=1}^{p} |\beta_i|^q.$$

There have been extensive studies towards understanding the performance of the best subset selection ($q = 0$), LASSO ($q = 1$) and ridge ($q = 2$), three widely known estimators from the LQLS family. This thesis aims at giving a unified view of LQLS for all the values of $q \in [0, \infty)$. In contrast to most existing works which obtain order-wise error bounds with loose constants, we derive asymptotically exact error formulas characterized through a series of fixed point equations. A delicate analysis of the fixed point equations enables us to gain fruitful insights into the statistical properties of LQLS across the entire spectrum of ℓ_q-regularization. Our work not only validates the scope of folklore understanding of ℓ_q-minimization, but also provides new insights into high-dimensional statistics as a whole. We will elaborate on our theoretical findings mainly from parameter estimation point of view. At the end of the thesis, we briefly mention bridge regression for variable selection and prediction.

We start by considering the parameter estimation problem and evaluate the performance of LQLS by characterizing the asymptotic mean square error (AMSE) $\lim_{n\to\infty} \frac{1}{p}\|\hat{\beta}(\lambda, q) - \beta\|_2^2$. The expression we derive for AMSE does not have explicit forms and hence is not useful in comparing LQLS for different values of q, or providing information in evaluating the effect of relative sample size $\frac{n}{p}$ or the sparsity level of β. To simplify the expression, we first perform the phase transition (PT) analysis, a widely accepted analysis diagram, of LQLS. Our results reveal some of the limitations and misleading features of the PT framework. To overcome these limitations, we propose the small-error analysis of LQLS. Our new analysis framework not only sheds light on the results of the phase transition analysis, but also describes when phase transition analysis is reliable, and presents a more accurate comparison among different ℓ_q-regularizations.

We then extend our low noise sensitivity analysis to linear models without sparsity structure. Our analysis, as a generalization of phase transition analysis, reveals a clear picture of bridge regression for estimating generic coefficients β. Moreover, by a simple transformation we connect our low-noise sensitivity framework to the classical asymptotic regime in which $n/p \to \infty$, and give some insightful implications beyond what classical asymptotic analysis of bridge regression can offer.

Furthermore, following the same idea of the new analysis framework, we are able to obtain an explicit characterization of AMSE in the form of second-order expansions under the large noise regime. The expansions provide us some intriguing messages. For example, ridge will outperform LASSO in terms of estimating sparse coefficients when the measurement noise is large.

Finally, we present a short analysis of LQLS, for the purpose of variable selection and prediction. We propose a two-stage variable selection technique based on the LQLS estimators, and describe its superiority and close connection to parameter estimation. For prediction, we illustrate the intricate relation between the tuning parameter selection for optimal in-sample prediction and optimal parameter estimation.

Table of Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank my Ph.D. advisors, Professors Yang Feng and Arian Maleki. It would have been impossible for me to complete the thesis without their patient guidance and continuous support. In particular, I offer my sincerest gratitude to Professor Yang Feng, for giving me intellectual freedom in my research and encouraging me to grow as a researcher. He has been a tremendous mentor and friend for me. I would also like to express my special appreciation to Professor Arian Maleki, for providing me a rewarding research experience on high-dimensional statistics. His advice on research and academic career has been invaluable.

I would like to thank my dissertation committee: Professors Tian Zheng, Peter Orbanz, and Zhezhen Jin, for their time and helpful suggestions. I would also like to acknowledge all the faculty and staff in the Department of Statistics at Columbia University. Thanks for all their amazing work to make the Ph.D. program and my Ph.D. journey so enjoyable. I am particularly grateful to Professors Zhiliang Ying and Richard Davis. They have been incredible academics and career mentors to me, and always willing to share their research and teaching insights. I am also greatly thankful to Dood Kalicharan and Anthony Cruz for their numerous assistance.

Additionally, I am deeply grateful to all my classmates and friends at Columbia. Their intelligence, hard work and personality is an invaluable source of inspiration and encouragement throughout my Ph.D. I would like to particularly thank my dear friend Yunxiao Chen. His support and help from daily life to academics plays a key role in the completion of my Ph.D. I am also thankful to Shuaiwen Wang for being a great collaborator and friend. Many thanks go to Jingjing Zou, Phyllis Wan, Resa

Juarez Maria, Rohit Patra, Timothy Jones, Jin Hyung Lee, Swupnil Sahai, Lu Meng, Xiaoou Li, Louis Mittel, Benjamin Reddy, Yang Kang, Yuanjun Gao, Lisha Qiu, Jing Zhang, Yuting Ma, Xuan Yang, Ying Liu, Tony Sit for their kindness and support, and the list goes on.

Last but not least, I would like to express my most sincere gratitude to my family. Thank my parents and brother for their infinite love and constant care. You mean the whole world to me. This thesis is for you.

To my family

Chapter 1

Introduction

1.1 Objective and organization

Consider the linear regression problem where the goal is to estimate the parameter vector $\beta \in \mathbb{R}^p$ from a set of n response variables $y \in \mathbb{R}^n$, under the model $y = X\beta + w$. This problem has been studied extensively in the last two centuries since Gauss and Legendre developed the least squares estimate of β. The instability or high variance of the least squares estimates led to the development of the regularized least squares. One of the most popular regularization classes is the ℓ_q-regularized least squares (LQLS), a.k.a. bridge regression (Frank and Friedman, 1993; Fu, 1998), given by the following optimization problem:

$$\hat{\beta}(\lambda, q) \in \arg\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \sum_{i=1}^{p} |\beta_i|^q. \tag{1.1}$$

where $q \in [0, \infty)$ and $\lambda > 0$ is a tuning parameter. LQLS has been extensively studied in the literature. In particular, one can prove the consistency of $\hat{\beta}(\lambda, q)$ under the classical asymptotic analysis (p fixed while $n \to \infty$) (Knight and Fu, 2000). However, this asymptotic regime becomes irrelevant for high-dimensional problems in which n is not much larger than p. Under this high-dimensional setting, if β does not have any specific structure, we do not expect any estimator to perform well.

One of the structures that has attracted attention in the last twenty years is the sparsity, that assumes only k of the elements of β are non-zero and the rest are zero. To understand the behavior of the estimators under structured linear model in high dimension, a new asymptotic framework has been proposed in which it is assumed that $X_{ij} \overset{i.i.d.}{\sim} N(0, 1/n), k, n, p \to \infty$ while $n/p \to \delta$ and $k/p \to \epsilon$, where δ and ϵ are fixed numbers (Donoho and Tanner, 2005b; Donoho et al., 2009; Amelunxen et al., 2014; El Karoui et al., 2013; Bradic and Chen, 2015).

One of the main notions that has been widely studied in this asymptotic framework, is the phase transition (Donoho and Tanner, 2005b; Donoho et al., 2009; Amelunxen et al., 2014; STOJNIC, 2009). Intuitively speaking, phase transition analysis assumes the error w equals zero and characterizes the value of δ above which an estimator converges to the true β (in certain sense that will be clarified in the following chapters). While there is always error in the response variables, it is believed that phase transition analysis provides reliable information when the errors are small. In this thesis, we start by studying the phase transition diagrams of LQLS for $q \in [0, \infty)$. Our analysis reveals several limitations of the phase transition analysis. We will clarify these limitations in Chapter 2. We then propose a higher-order analysis of LQLS in the small-error regime. As will be explained in Chapter 2, our new framework sheds light on the peculiar behavior of the phase transition diagrams, and explains when we can rely on the results of phase transition analysis in practice.

The sparsity assumption of β can be easily violated in many applications. A more realistic replacement is to assume β is approximately sparse, i.e., some elements of β are very small. Under such a setting, all the asymptotic results we derived for sparse coefficients may not hold any more. However, we will demonstrate in Chapter 3 that, the limitations of phase transition analysis remain. We then perform a low-noise sensitivity analysis as a generalization of phase transition scheme to better evaluate and compare different LQLS estimators. Moreover, by a simple transformation we connect our low-noise sensitivity framework to the classical asymptotic regime in which

$n/p \to \infty$ and characterize how and when ℓ_q regularization techniques offer improvements over ordinary least squares, and which regularizer gives the most improvement when the sample size is large.

The small-error analysis enables us to have a more accurate evaluation and comparison for different LQLS estimators when the measurement noise w is small. However, the results can not carry over to settings with large noises. Motivated by this concern, we further perform a second-order noise sensitivity analysis under large-error regime. Our analysis discovers an intriguing phenomenon regarding the parameter estimation performance of LQLS: ridge is optimal among all the LQLS estimators. This implies that in low signal-to-noise ratios, sparsity promoting regularization methods like LASSO and best subset selection are inferior to ridge, even though the estimand is sparse. We present a thorough comparison of LQLS for every value of $q \geq 0$ in Chapter 4.

If the primary interest lies on variable selection or prediction instead of parameter estimation, how would the performance of LQLS estimators change? In Chapter 5 we scratch the surface of these two directions. For the former, we first propose a two-stage variable selection technique with LQLS estimators used in the first stage. It will be shown that the two-stage LASSO, one example of the proposed approach, outperforms LASSO. More importantly, we establish the equivalence between the variable selection comparison and parameter estimation comparison among different LQLS's. For the latter, we present some preliminary results regarding the tuning parameter selection for optimal prediction.

1.2 The asymptotic framework

The main goal of this section is to formally introduce the high-dimensional asymptotic framework under which we study LQLS throughout the thesis. We may write vectors and matrices as $\beta(p), X(p), w(p)$ to emphasize the dependence on the dimension of β.

Similarly, we may use $\hat{\beta}(\lambda, q, p)$ as a substitute for $\hat{\beta}(\lambda, q)$. We first define a specific type of a sequence known as a converging sequence. Our definition is borrowed from other papers (Donoho et al., 2011; Bayati and Montanari, 2011, 2012) with some minor modifications. Recall we have the linear regression model: $y(p) = X(p)\beta(p) + w(p)$.

Definition 1.2.1. *A sequence instances* $\{\beta(p), X(p), w(p)\}$ *is called a converging sequence if the following conditions hold:*

1. $n/p \to \delta \in (0, \infty)$, *as* $n \to \infty$.

2. *The empirical distribution*[1] *of* $\beta(p) \in \mathbb{R}^p$ *converges weakly to a probability measure* f_β *with bounded second moment. Further,* $\frac{1}{p}\|\beta(p)\|_2^2$ *converges to the second moment of* f_β.

3. *The empirical distribution of* $w(p) \in \mathbb{R}^n$ *converges weakly to a zero mean distribution with variance* σ_w^2. *And,* $\frac{1}{n}\|w(p)\|_2^2 \to \sigma_w^2$.

4. *The elements of* $X(p)$ *are iid with distribution* $N(0, 1/n)$.

For each of the problem instances in a converging sequence, we solve the LQLS problem (1.1) and obtain $\hat{\beta}(\lambda, q, p)$ as the estimator. The interest is to evaluate the accuracy of this estimator. For different purposes such as parameter estimation and variable selection, we will define different quantities to measure the performance of LQLS in the following chapters.

---

[1] It is the distribution that puts a point mass $1/p$ at each of the p elements of the vector.

Chapter 2

Overcoming the limitations of phase transition via a second-order low noise sensitivity analysis

2.1 Limitations of the phase transition and our solution

In this section, we intuitively describe the results of phase transition analysis, its limitations, and our new framework. Consider the class of LQLS estimators and suppose that we would like to compare the performance of these estimators through the phase transition diagrams. For the purpose of this section, we assume that the vector β has only k non-zero elements, where $k/p \to \epsilon$ with $\epsilon \in (0, 1)$. Since phase transition analysis is concerned with $w = 0$ setting, it considers $\lim_{\lambda \to 0} \hat{\beta}(\lambda, q)$ which is equivalent to the following estimator:

$$\begin{aligned}&\arg\min_{\beta} \|\beta\|_q^q, \\ &\text{subject to } y = X\beta. \end{aligned} \tag{2.1}$$

Below we informally state the results of the phase transition analysis. We will formalize the statement and describe in details the conditions under which this result holds in the next section.

Informal Result 1. For a given $\epsilon > 0$ and $q \in [1, \infty)$, there exists a number $M_q(\epsilon)$ such that as $p \to \infty$, if $\delta \geq M_q(\epsilon) + \gamma$ ($\gamma > 0$ is an arbitrary number), then (2.1) succeeds in recovering β, while if $\delta \leq M_q(\epsilon) - \gamma$, (2.1) fails.[1]

The curve $\delta = M_q(\epsilon)$ is called the phase transition curve of (2.1). While the phase transition curves can be obtained with different techniques, such as statistical dimension framework proposed in Amelunxen et al. (2014), we will derive them as a simple byproduct of our main results in the next section. We will show that $M_q(\epsilon)$ is given by the following formula:

$$M_q(\epsilon) = \begin{cases} 1 \ \text{ if } q > 1, \\ \inf_{\chi \geq 0} (1-\epsilon)\mathbb{E}\eta_1^2(Z;\chi) + \epsilon(1+\chi^2) \ \text{ if } q = 1, \\ \epsilon \ \text{ if } 1 > q \geq 0, \end{cases} \tag{2.2}$$

where $\eta_1(u;\chi) = (|u| - \chi)_+ \text{sign}(u)$ denotes the soft thresholding function and $Z \sim N(0,1)$. Before we proceed further let us mention some of the properties of $M_1(\epsilon)$ that will be useful in our later discussions.

Lemma 2.1.1. *$M_1(\epsilon)$ satisfies the following properties:*

(i) $M_1(\epsilon)$ *is an increasing function of* ϵ.

(ii) $\lim_{\epsilon \to 0} M_1(\epsilon) = 0$.

(iii) $\lim_{\epsilon \to 1} M_1(\epsilon) = 1$.

(iv) $M_1(\epsilon) > \epsilon$, *for* $\epsilon \in (0, 1)$.

[1]Different notions of success have been studied in the phase transition analysis. We will mention one notion later in this thesis.

Proof. Define $F(\chi, \epsilon) \triangleq (1-\epsilon)\mathbb{E}\eta_1^2(Z;\chi)+\epsilon(1+\chi^2)$. It is straightforward to verify that $F(\chi, \epsilon)$, as a function of χ over $[0, \infty)$, is strongly convex and has a unique minimizer. Let $\chi^*(\epsilon)$ be the minimizer. We write it as $\chi^*(\epsilon)$ to emphasize its dependence on ϵ. By employing the chain rule we have

$$\begin{aligned}\frac{dM_1(\epsilon)}{d\epsilon} &= \frac{\partial F(\chi^*(\epsilon), \epsilon)}{\partial \epsilon} + \frac{\partial F(\chi^*(\epsilon), \epsilon)}{\partial \chi} \cdot \frac{d\chi^*(\epsilon)}{d\epsilon} = \frac{\partial F(\chi^*(\epsilon), \epsilon)}{\partial \epsilon} \\ &= 1 + (\chi^*(\epsilon))^2 - \mathbb{E}\eta_1^2(Z; \chi^*(\epsilon)) > 1 + (\chi^*(\epsilon))^2 - \mathbb{E}|Z|^2 \\ &= (\chi^*(\epsilon))^2 > 0,\end{aligned}$$

which completes the proof of part (i). To prove (ii) note that

$$\begin{aligned} 0 &\leq \lim_{\epsilon \to 0} \min_{\chi \geq 0} (1-\epsilon)\mathbb{E}\eta_1^2(Z;\chi) + \epsilon(1+\chi^2) \\ &\leq \lim_{\epsilon \to 0} (1-\epsilon)\mathbb{E}\eta_1^2(Z; \log(1/\epsilon)) + \epsilon(1 + \log^2(1/\epsilon)) \\ &= \lim_{\epsilon \to 0} 2(1-\epsilon) \int_{\log(1/\epsilon)}^{\infty} (z - \log(1/\epsilon))^2 \phi(z) dz \\ &= \lim_{\epsilon \to 0} 2(1-\epsilon) \int_0^{\infty} z^2 \phi(z + \log(1/\epsilon)) dz \\ &\leq \lim_{\epsilon \to 0} 2(1-\epsilon) \mathrm{e}^{-\frac{\log^2(1/\epsilon)}{2}} \int_0^{\infty} z^2 \phi(z) dz = 0, \end{aligned}$$

where $\phi(\cdot)$ is the density function of standard normal. Regarding the proof of part (iii), first note that as $\epsilon \to 1$, $\chi^*(\epsilon) \to 0$. Otherwise suppose $\chi^*(\epsilon) \to \chi_0 > 0$ (taking a convergent subsequence if necessary). Since $\mathbb{E}\eta_1^2(Z; \chi^*(\epsilon)) \leq \mathbb{E}|Z|^2 = 1$, we obtain

$$\lim_{\epsilon \to 1} F(\chi^*(\epsilon), \epsilon) = 1 + \chi_0^2 > 1.$$

On the other hand, it is clear that

$$\lim_{\epsilon \to 1} F(\chi^*(\epsilon), \epsilon) \leq \lim_{\epsilon \to 1} F(0, \epsilon) = 1.$$

A contradiction arises. Hence the fact $\chi^*(\epsilon) \to 0$ as $\epsilon \to 1$ leads directly to

$$\lim_{\epsilon \to 1} M_1(\epsilon) = \lim_{\epsilon \to 1} F(\chi^*(\epsilon), \epsilon) = 1.$$

Part (iv) is clear from the definition of $M_1(\epsilon)$. □

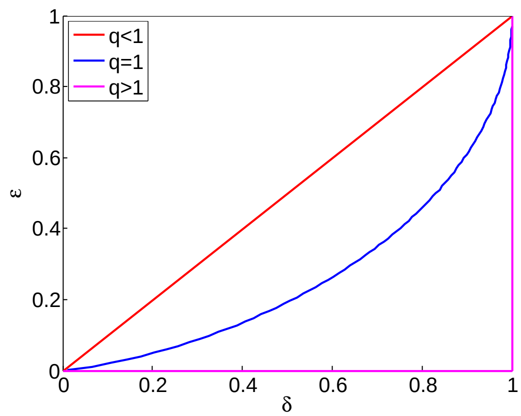


Figure 2.1: Phase transition curves of LQLS for (i) $q < 1$: The red curve denotes the phase transition of LQLS for any $q \in [0, 1)$. (ii) $q = 1$: The blue curve exhibits the phase transition of LASSO. Below this curve LASSO can "successfully" recover β. (iii) $q > 1$: The magenta curve represents the phase transition of LQLS for any $q > 1$. This figure is based on Informal Result 1 and will be carefully defined and derived in Chapter 2.2.

Figure 2.1 shows $M_q(\epsilon)$ for different values of q. We observe several peculiar features: (i) As is clear from both Lemma 2.1.1 and Figure 2.1, $q = 1$ requires much fewer observations than all the values of $q > 1$ and much more observations than all $q \in [0, 1)$ for successful recovery of β. (ii) The values of the non-zero elements of β do not have any effect on the phase transition curves. In fact, even the sparsity level does not have any effect on the phase transition for $q > 1$. (iii) For every $q > 1$, the phase transition of (2.1) happens at exactly the same value. So does every value of q belonging to $[0, 1)$.

These features raise the following question: how much and to what extent are these phase transition results useful in applications, where at least small amount of error is present in the response variables? For instance, intuitively speaking, we do not expect to see much difference between the performance of LQLS for $q = 1.01$ and $q = 1$. However, according to the phase transition analysis, $q = 1$ outperforms

$q = 1.01$ by a wide margin. In fact the performance of LQLS for $q = 1.01$ seems to be closer to that of $q = 2$ than $q = 1$. The same reasoning goes for $q = 1$ and $q = 0.99$. Also, in contrast to the phase transition implication, we may not expect LQLS to perform the same for β with different values of non-zero elements. The main goal of this chapter is to present a new analysis that will shed light on the misleading features of the phase transition analysis. It will also clarify when and under what conditions the phase transition analysis is reliable for practical guidance.

In our new framework, the variance σ_w^2 of the error w is assumed to be small. We consider (1.1) with the optimal value of λ for which the asymptotic mean square error, i.e., $\lim_{p\to\infty} \frac{\|\hat{\beta}(\lambda,q)-\beta\|_2^2}{p}$, is minimized. We first obtain the formula for the asymptotic mean square error (AMSE) characterized through a series of non-linear equations. Since σ_w is assumed small, we then derive the second-order asymptotic expansions for AMSE as $\sigma_w \to 0$. As we will describe later, the phase transition of LQLS for different values of q can be obtained from the first dominant term in the expansion. More importantly, we will show that the second dominant term is capable of evaluating the importance of the phase transition analysis for practical situations and also provides a much more accurate analysis of different bridge estimators. Here is one of our main results, presented informally to clarify our claims. All the technical conditions will be determined in Chapter 2.2.

Informal Result 2. If $\lambda_{*,q}$ denotes the optimal value of λ, then for any $q \in (1, 2)$, $\delta > 1$, and $\epsilon < 1$

$$\lim_{p\to\infty} \frac{1}{p}\|\hat{\beta}(\lambda_{*,q}, q) - \beta\|_2^2 = \frac{\sigma_w^2}{1 - 1/\delta} - \sigma_w^{2q} \frac{\delta^{q+1}(1-\epsilon)^2(\mathbb{E}|Z|^q)^2}{(\delta-1)^{q+1}\epsilon\mathbb{E}|G|^{2q-2}} + o(\sigma_w^{2q}),$$

where $Z \sim N(0, 1)$ and G is a random variable whose distribution is specified by the non-zero elements of β. We will clarify this in the next section. Finally, the limit notation we have used above is the almost sure limit.

As we will discuss in Chapter 2.2, the first term $\frac{\sigma_w^2}{1-1/\delta}$ determines the phase transition. Moreover, we have further derived the second dominant term in the expansion of the asymptotic mean square error. This term enables us to clarify some of the confusing features of the phase transitions. Here are some important features of this term: (i) It is negative. Hence, the AMSE that is predicted by the first term (and phase transition analysis) is overestimated specially when q is close to 1. (ii) Fixing q, the magnitude of the second dominant term grows as ϵ decreases. Hence, for small values of σ_w all values of $1 < q < 2$ benefit from the sparsity of β. Also, smaller values of q seem to benefit more. (iii) Fixing ϵ and δ, the power of σ_w decreases as q decreases. This makes the absolute value of the second dominant term bigger. As q decreases to one, the order of the second dominant term gets closer to that of the first dominant term and thus the predictions of phase transition analysis become less accurate. We will present a more detailed discussion of the second order term in Chapter 2.2. To show some more interesting features of our approach, we also informally state a result we prove for LASSO.

Informal Result 3. Suppose that the non-zero elements of β are all larger than a fixed number μ. If $\lambda_{*,q}$ denotes the value of λ that leads to the smallest AMSE, and if $\delta > M_1(\epsilon)$, then for $q = 1$

$$\lim_{p\to\infty} \frac{1}{p}\|\hat{\beta}(\lambda_{*,q}, q) - \beta\|_2^2 = \frac{\delta M_1(\epsilon)\sigma_w^2}{\delta - M_1(\epsilon)} + O(\exp(-\tilde{\mu}/\sigma_w^2)), \tag{2.3}$$

where $\tilde{\mu}$ is a constant that depends on μ.

As can be seen here, compared to other values of $1 < q < 2$, $q = 1$ has smaller first order term (according to Lemma 2.1.1), but much smaller (in magnitude) second order term. The first implication of this result is that the first dominant term provides an accurate approximation of AMSE. Hence, phase transition analysis in this case is reliable even if small amount of noise is present; that is one of the main reasons why the theoretically derived phase transition curve matches the empirical one for LASSO. Furthermore, note that in order to obtain Informal Result 3, we have made certain assumption about the non-zero components of β. As will be shown in Chapter 2.2, any violation of this assumption has major impact on the second dominant term.

2.2 A second-order low noise sensitivity analysis

2.2.1 Characterization of asymptotic mean square error

We define the asymptotic mean square error of LQLS estimators below to measure their accuracy.

Definition 2.2.1. *Let $\hat{\beta}(\lambda, q, p)$ be the sequence of solutions of LQLS for the converging sequence of instances $\{\beta(p), X(p), w(p)\}$. The asymptotic mean square error is defined as the almost sure limit of*

$$\text{AMSE}(\lambda, q, \sigma_w) \triangleq \lim_{p\to\infty} \frac{1}{p} \sum_{i=1}^{p} |\hat{\beta}_i(\lambda, q, p) - \beta_i(p)|^2,$$

where the subscript i is used to denote the ith component of a vector.

Note that we have suppressed δ and f_β in the notation of AMSE for simplicity, despite the fact that the asymptotic mean square error depends on them as well. In the above definition, we have assumed that the almost sure limit exists. Under the current asymptotic setting introduced in Chapter 1.2, the existence of AMSE can be proved. We state the results for $q \geq 1$ and $0 \leq q < 1$, respectively.

Theorem 2.2.2. *Consider a converging sequence $\{\beta(p), X(p), w(p)\}$. For any given $q \in [1, \infty)$, suppose that $\hat{\beta}(\lambda, q, p)$ is the solution of LQLS defined in (1.1). Then for any pseudo-Lipschitz function[2] $\psi : \mathbb{R}^2 \to \mathbb{R}$, almost surely*

$$\lim_{p\to\infty} \frac{1}{p} \sum_{i=1}^{p} \psi\left(\hat{\beta}_i(\lambda, q, p), \beta_i(p)\right) = \mathbb{E}_{B,Z}[\psi(\eta_q(B + \bar{\sigma} Z; \bar{\chi}\bar{\sigma}^{2-q}), B)], \quad (2.4)$$

[2] A function $\psi : \mathbb{R}^2 \to \mathbb{R}$ is pseudo-Lipschitz of order k if there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^2$, we have $|\psi(x) - \psi(y)| \leq L(1 + \|x\|_2^{k-1} + \|y\|_2^{k-1})\|x - y\|_2$. We consider pseudo-Lipschitz functions with order 2 in this thesis.

where B and Z are two independent random variables with distributions f_β and $N(0,1)$, respectively; the expectation $\mathbb{E}_{B,Z}(\cdot)$ is taken with respect to both B and Z; $\eta_q(\cdot;\cdot)$ is the proximal operator for the function $\|\cdot\|_q^q$[3]; and $(\bar{\sigma},\bar{\chi})$ is the unique pair satisfying the following equations:

$$\begin{aligned} \bar{\sigma}^2 &= \sigma_\omega^2 + \frac{1}{\delta}\mathbb{E}_{B,Z}[(\eta_q(B+\bar{\sigma}Z;\bar{\chi}\bar{\sigma}^{2-q}) - B)^2], && (2.5)\\ \lambda &= \bar{\chi}\bar{\sigma}^{2-q}\Big(1-\frac{1}{\delta}\mathbb{E}_{B,Z}[\eta_q'(B+\bar{\sigma}Z;\bar{\chi}\bar{\sigma}^{2-q})]\Big), && (2.6)\end{aligned}$$

where $\eta_q'(\cdot;\cdot)$ denotes the derivative of η_q with respect to its first argument.

The result for $q=1$ has been proved in Bayati and Montanari (2012). The key ideas of the proof for generalizing to $q \in (1,\infty)$ are similar to those of Bayati and Montanari (2012). We describe the main proof steps in Chapter 2.5.4.

Theorem 2.2.3. *Consider a special converging sequence $\{\beta(p), X(p), w(p)\}$ where the elements of $\beta(p)$ are iid from f_β and the components of $w(p)$ are iid from a zero-mean distribution with variance σ_w^2. For any given $q \in [0,1)$, suppose there exists a random variable S such that $|\hat{\beta}_1(\lambda,q)| < S$ for every value of p and $\mathbb{E}|S|^2 < \infty$, then under the assumptions of replica method (Rangan et al., 2012), almost surely*

$$\mathrm{AMSE}(\lambda,q,\sigma_w) = \mathbb{E}_{B,Z}[\eta_q(B+\bar{\sigma}Z;\bar{\chi}\bar{\sigma}^{2-q}) - B]^2, \quad (2.7)$$

where $B, Z, \bar{\sigma}, \bar{\chi}, \eta_q(\cdot;\cdot)$ are the same as in Theorem 2.2.2.

The proof, will be shown in Chapter 2.5.5, is a direct application of replica claim in Rangan et al. (2012). The replica method is a widely accepted and powerful heuristic method in statistical physics for analyzing large disordered systems (Mézard et al., 1987). It has been adapted to attack theoretical problems in other fields like compressed sensing (Rangan et al., 2012) and network analysis (Decelle et al., 2011). Some of its important predictions have been rigorously proved (Bayati and

[3]Proximal operator of $\|\cdot\|_q^q$ is defined as $\eta_q(u;\chi) \triangleq \arg\min_z \frac{1}{2}(u-z)^2 + \chi|z|^q$. For further information on these functions, refer to Chapter 2.5.3.

Montanari, 2012; Mossel et al., 2015; Massoulié, 2014; Mossel et al., 2013). Theorem 2.2.3 relies on the replica assumptions. The validation of its full rigorousness remains an open problem. Nevertheless, we are able to design an approximate message passing algorithm for solving (1.1) with $0 \leq q < 1$, and rigorously show the asymptotic mean square error of the output from the algorithm takes the same expression as in Theorem 2.2.3. Refer to Zheng et al. (2017) for the details.

Theorems 2.2.2 and 2.2.3 provide the first step in our analysis of LQLS. We first calculate $\bar{\sigma}$ and $\bar{\chi}$ from (2.5) and (2.6). Then, incorporating $\bar{\sigma}$ and $\bar{\chi}$ in (2.7) yields the asymptotic mean square error. Given the distribution f_β, the variance of the error σ_w^2, the number of response variables (normalized by the number of predictors) δ, and the regularization parameter λ, it is straightforward to write a computer program to find the solution of (2.5) and (2.6) and then compute the value of AMSE. However, it is needless to say that this approach does not shed much light on the performance of bridge regression estimates, since there are many factors involved in the computation and each affects the result in a non-trivial fashion. In the rest of this chapter, we would like to perform an analytical study on the solution of (2.5) and (2.6) and obtain an explicit characterization of AMSE in the small-error regime.

2.2.2 Optimal tuning of λ

The performance of LQLS, as defined in (1.1), depends on the tuning parameter λ. We consider the value of λ that gives the minimum AMSE. Let $\lambda_{*,q}$ denote the value of λ that minimizes AMSE given in (2.7). Then LQLS is solved with this specific value of λ, i.e.,

$$\hat{\beta}(\lambda_{*,q}, q, p) \in \arg\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda_{*,q}\|\beta\|_q^q. \tag{2.8}$$

Note that this is the best performance that LQLS can achieve in terms of the AMSE. Theorems 2.2.2 and 2.2.3 enable us to evaluate this optimal AMSE of LQLS for every $q \in [0, \infty)$. The key step is to compute the solution of (2.5) and (2.6) with $\lambda = \lambda_{*,q}$.

Since $\lambda_{*,q}$ has to be chosen optimally, it seemingly causes an extra complication for our analysis. However, as we show in the following corollary, the study of Equations (2.5) and (2.6) can be simplified to some extent.

Corollary 2.2.4. *Suppose that* $\hat{\beta}(\lambda_{*,q}, q, p)$ *is the solution of LQLS defined in* (2.8)*, and the conditions in Theorems 2.2.2 and 2.2.3 hold for* $q \geq 1$ *and* $0 \leq q < 1$*, respectively. Then for any* $q \in [0, \infty)$*,*

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \min_{\chi \geq 0} \mathbb{E}_{B,Z}(\eta_q(B + \bar{\sigma}Z; \chi) - B)^2, \tag{2.9}$$

where B *and* Z *are two independent random variables with distributions* f_β *and* $N(0,1)$*, respectively; and* $\bar{\sigma}$ *is the unique solution of the following equation:*

$$\bar{\sigma}^2 = \sigma_\omega^2 + \frac{1}{\delta} \min_{\chi \geq 0} \mathbb{E}_{B,Z}[(\eta_q(B + \bar{\sigma}Z; \chi) - B)^2]. \tag{2.10}$$

The proof of Corollary 2.2.4 is shown in Chapter 2.5.6. Corollary 2.2.4 enables us to focus the analysis on a single equation (2.10), rather than two equations (2.5) and (2.6). The results we will present are mainly based on investigating the solution of (2.10).

2.2.3 Second-order expansions of asymptotic mean square error

Since we have been focused on the sparsity structure of β, in the rest of this chapter we assume that the distribution, to which the empirical distribution of $\beta \in \mathbb{R}^p$ converges, has the form

$$f_\beta(b) = (1 - \epsilon)\delta_0(b) + \epsilon g(b),$$

where $\delta_0(\cdot)$ denotes a point mass at zero, and $g(\cdot)$ is a generic distribution that does not have any point mass at 0. Here, the mixture proportion $\epsilon \in (0,1)$ is a fixed number that represents the sparsity level of β. The smaller ϵ is, the sparser β will be. The distribution $g(b)$ specifies the values of non-zero components of β. We will use G

to denote a random variable having such a distribution. We also use Z to represent a standard normal. Since our results and proof techniques look very different for $0 \leq q < 1, q = 1, q > 1$, we study these cases separately.

2.2.3.1 Results for $q > 1$

Our first result is concerned with the optimal AMSE of LQLS for $q > 1$, when the number of response variables is larger than the number of predictors p, i.e., $\delta > 1$.

Theorem 2.2.5. *Suppose* $\epsilon \in (0,1), \delta > 1$. *For* $1 < q < 2$, *if* $\mathbb{P}(|G| \leq t) = O(t)$ *(as* $t \to 0$*) and* $\mathbb{E}|G|^2 < \infty$, *we have*

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1-1/\delta} - \frac{\delta^{q+1}(1-\epsilon)^2(\mathbb{E}|Z|^q)^2}{(\delta-1)^{q+1}\epsilon\mathbb{E}|G|^{2q-2}}\sigma_w^{2q} + o(\sigma_w^{2q}). \quad (2.11)$$

For $q = 2$, *if* $\mathbb{E}|G|^2 < \infty$ *we have*

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1-1/\delta} - \frac{\delta^3\sigma_w^4}{(\delta-1)^3\epsilon\mathbb{E}|G|^2} + o(\sigma_w^4).$$

For $q > 2$, *if* $\mathbb{E}|G|^{2q-2} < \infty$ *then*

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1-1/\delta} - \frac{\delta^3\epsilon(q-1)^2(\mathbb{E}|G|^{q-2})^2\sigma_w^4}{(\delta-1)^3\mathbb{E}|G|^{2q-2}} + o(\sigma_w^4).$$

The proof of the result is presented in Chapter 2.5.7. There are several interesting features of this result that we would like to discuss: (i) The second dominant term of AMSE is negative. This means that the actual AMSE is smaller than the one predicted by the first order term, especially for smaller values of q. (ii) Neither the sparsity level nor the distribution of the non-zero components of β appear in the first dominant term, i.e. $\frac{\sigma_w^2}{1-1/\delta}$. As we will discuss later in this section, the first dominant term is the one that specifies the phase transition curve. Hence, these calculations show a peculiar feature of phase transition analysis we discussed in Chapter 2.1, that the phase transition of $q \in (1, \infty)$ is neither affected by non-zero components of β or the sparsity level. However, we see that both factors come into play in the second dominant term. (iii) For the fully dense coefficient, i.e. $\epsilon = 1$, (2.11) may imply that for $1 < q < 2$,

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1 - 1/\delta} + o(\sigma_w^{2q}).$$

Hence, we require a different analysis to obtain the second dominant term (with different orders). We present a full discussion for non-sparse coefficients in Chapter 3. (iv) For $\epsilon < 1$, the choice of $q \in (1, \infty)$ does not affect the first dominant term. That is the reason why all the values of $q \in (1, \infty)$ share the same phase transition curve. However, the value of q has a major impact on the second dominant term. In particular, as q approaches 1, the order of the second dominant term in terms of σ_w gets closer to that of the first dominant term. This means that in any practical setting, phase transition analysis may lead to misleading conclusions. Specifically, in contrast to the conclusion from phase transition analysis that $q \in (1, \infty)$ have the same performance, the second order expansion enables us to conclude that, for $q \in (1, 2]$ the closer to 1 the value of q is, the better its performance will be. And interestingly such monotonicity does not hold any more beyond $q = 2$. Our next theorem discusses the AMSE when $\delta < 1$.

Theorem 2.2.6. *Suppose $\mathbb{E}|G|^2 < \infty$, then for $q > 1$ and $\delta < 1$,*

$$\text{AMSE}(\lambda_{*,q}, q, 0) > 0. \tag{2.12}$$

The proof of this theorem is presented in Chapter 2.5.8. Theorems 2.2.5 and 2.2.6 together show a notion of phase transition. For $\delta > 1$, as $\sigma_w \to 0$, AMSE $= O(\sigma_w^2)$, and hence it will go to zero, while AMSE $\nrightarrow 0$ for $\delta < 1$. In fact, the phase transition curve $\delta = 1$ can be derived from the first dominant term in the expansion of AMSE. If $\delta = 1$, the first dominant term is infinity and there will be no successful recovery, while it becomes zero when $\sigma_w = 0$ if $\delta > 1$. A more rigorous justification can be found in the proof of Theorems 2.2.5 and 2.2.6. Therefore, we may conclude that the first order term contains the phase transition information. Moreover, the derived second order term offers us additional important information regarding the accuracy of the phase transition analysis. To provide a comprehensive understanding of these two terms, in Chapter 2.3 we will evaluate the accuracy of first and second order approximations to AMSE through numerical studies.

2.2.3.2 Results for $q = 1$

So far we have studied the case $q > 1$. In this section, we study $q = 1$, a.k.a. LASSO. In Theorems 2.2.5 and 2.2.6, we have characterized the behavior of LQLS with $q \in (1, \infty)$ for a general class of G. It turns out that the distribution of G has a more serious impact on the second dominant term of AMSE for LASSO. We thus analyze it in two different settings. Our first theorem considers the distributions that do not have any mass around zero.

Theorem 2.2.7. *Suppose* $\mathbb{P}(|G| > \mu) = 1$ *with* μ *being a positive constant and* $\mathbb{E}|G|^2 < \infty$, *then for* $\delta > M_1(\epsilon)$[4]

$$\text{AMSE}(\lambda_{*,1}, 1, \sigma_w) = \frac{\delta M_1(\epsilon)}{\delta - M_1(\epsilon)} \sigma_w^2 - |o\big(e^{\frac{(M_1(\epsilon)-\delta)\tilde{\mu}^2}{2\delta\sigma_w^2}}\big)|, \tag{2.13}$$

where $\tilde{\mu}$ *is any positive constant smaller than* μ.

The proof of Theorem 2.2.7 is given in Chapter 2.5.2. Different from the case for LQLS with $q \in (1, \infty)$, we have not derived the exact analytical expression of the second dominant term for LASSO. However, since it is exponentially small, the first order term (or phase transition analysis) is sufficient for evaluating the performance of LASSO in the small-error regime. This will be further confirmed by the numerical studies in Chapter 2.3. Below is our result for the distributions of G that have more mass around zero.

---

[4]Recall $M_1(\epsilon) = \inf_{\chi \geq 0}(1-\epsilon)\mathbb{E}\eta_1^2(Z;\chi) + \epsilon(1+\chi^2)$ with $Z \sim N(0,1)$.

Theorem 2.2.8. *Suppose that* $\mathbb{P}(|G| \leq t) = \Theta(t^\ell)$ *(as* $t \to 0$*) with* $\ell > 0$ *and* $\mathbb{E}|G|^2 < \infty$*, then for* $\delta > M_1(\epsilon)$*,*

$$\begin{aligned} -|\Theta(\sigma_w^{\ell+2})| &\gtrsim \text{AMSE}(\lambda_{*,1}, 1, \sigma_w) - \frac{\delta M_1(\epsilon)}{\delta - M_1(\epsilon)}\sigma_w^2 \\ &\gtrsim -|\Theta(\sigma_w^{\ell+2})| \cdot \big(\underbrace{\log\log\ldots\log}_{m \text{ times}}(1/\sigma_w)\big)^{\ell/2}, \end{aligned}$$

where m *is an arbitrary but finite natural number.*

The proof of this theorem can be found in Chapter 2.5.9. It is important to notice the difference between Theorems 2.2.7 and 2.2.8. The first point we would like to emphasize is that the first dominant terms are the same in both cases. The second dominant terms are different though. Similar to LQLS for $q > 1$, the second dominant terms are in fact negative. Hence, the actual AMSE will be smaller than the one predicted by the first dominant term. Furthermore, note that the magnitude of the second dominant term in Theorem 2.2.8 is much larger than that in Theorem 2.2.7. This seems intuitive. LASSO tends to shrink the parameter coefficients towards zero, and hence, if the true β has more mass around zero, the AMSE will be smaller. The more mass the distribution of G has around zero, the better the second order term will be. Our next theorem discusses what happens if $\delta < M_1(\epsilon)$.

Theorem 2.2.9. *Suppose that* $\mathbb{E}|G|^2 < \infty$*. Then for* $\delta < M_1(\epsilon)$*,*

$$\text{AMSE}(\lambda_{*,1}, 1, 0) > 0. \tag{2.14}$$

The proof is presented in Chapter 2.5.10. Similarly as we discussed in Chapter 2.2.3.1, Theorems 2.2.7, 2.2.8 and 2.2.9 together imply the phase transition curve of LASSO. Such information can be obtained from the first dominant term in the expansion of AMSE as well.

2.2.3.3 Results for $0 \leq q < 1$

Theorem 2.2.10. *Suppose* $\mathbb{P}(|G| > \mu) = 1$ *with* μ *being a positive constant and* $\mathbb{E}|G|^2 < \infty$*, then for* $0 < q < 1, \delta > \epsilon$,

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\epsilon\delta}{\delta - \epsilon}\sigma_w^2 + \frac{\epsilon(4-4q)^{2-q}q^2\mathbb{E}|G|^{2q-2}\delta^{3-q}}{c_q^{4-2q}(\delta-\epsilon)^{3-q}}(\log 1/\sigma_w)^{2-q}\sigma_w^{4-2q} + o((\log 1/\sigma_w)^{2-q}\sigma_w^{4-2q}),$$

where $c_q = [2(1-q)]^{\frac{1}{2-q}} + q[2(1-q)]^{\frac{q-1}{2-q}}$.

Theorem 2.2.11. *Suppose* $\mathbb{P}(|G| > \mu) = 1$ *with* $\mu = \sup_v\{v : \mathbb{P}(|G| > v) = 1\} > 0$ *and* $\mathbb{E}|G|^2 < \infty$*, then for* $\delta > \epsilon$,

$$\text{AMSE}(\lambda_{*,0}, 0, \sigma_w) = \frac{\epsilon\delta}{\delta - \epsilon}\sigma_w^2 + o(e^{\frac{-\tilde{\mu}^2}{2\sigma_w^2}}),$$

where $\tilde{\mu}$ *is any constant that is smaller than* $\frac{\mu}{2}\sqrt{\frac{\delta-\epsilon}{\delta}}$.

The proof of Theorems 2.2.10 and 2.2.11 can be found in Chapters 2.5.11 and 2.5.12, respectively. There are again several interesting features of the above two results that we would like to emphasize. (i) The first dominant term in the expansion of AMSE is the same for $0 \leq q < 1$ and is smaller than that for LASSO. This is consistent with the phase transition analysis we presented in Chapter 2.1. (ii) The second dominant term is positive for $0 < q < 1$. In other words, the AMSE that is predicted by the first dominant term is smaller than the actual AMSE. Also, ignoring the logarithmic factors, the second dominant terms is proportional to σ_w^{4-2q} for $0 < q < 1$ and is exponentially small for $q = 0$. Hence, ℓ_0-regularization has the best performance, and as q gets closer to 0, the performance gets better. (iii) The above two theorems also reveal the impact of the distribution of the non-zero components of β, i.e. G, on $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$. Unlike the phase transition analysis, to obtain the above results, we have made some assumptions on G. We believe that the distribution of G has a major effect on the second dominant term for $q = 0$. We leave a delicate analysis like Theorem 2.2.8 as a future work.

Theorem 2.2.12. *Suppose that* $\mathbb{E}|G|^2 < \infty$. *Then for* $q \in [0, 1), \delta < \epsilon$,

$$\text{AMSE}(\lambda_{*,q}, q, 0) > 0.$$

The proof is presented in Chapter 2.5.13. Theorems 2.2.10, 2.2.11 and 2.2.12 together fully characterize the phase transition diagram for $0 \leq q < 1$.

Remark: For $0 \leq q < 1$, the optimization problem (1.1) is non-convex and computing the global optimum $\hat{\beta}(\lambda, q)$ is NP-hard. Hence in addition to Theorems 2.2.10, 2.2.11 and 2.2.12 concerned with $\hat{\beta}(\lambda, q)$, an interesting and important problem is to characterize the performance of some practical algorithms aimed for solving (1.1). Towards that goal, we have proposed an approximate message passing algorithm for (1.1) and given a comprehensive analysis of its statistical properties under different initializations. Refer to Zheng et al. (2017) for all the relevant results.

2.3 Numerical experiments

The analysis of AMSE we presented in Chapter 2.2.3 is performed as $\sigma_w \to 0$. For such asymptotic analysis, it would be interesting to check the approximation accuracy of the first and second order expansions of AMSE over a reasonable range of σ_w. Towards this goal, this section performs several numerical studies to (i) evaluate the accuracy of the first and second order expansions discussed in Chapter 2.2.3, (ii) discover situations in which the first order approximation is not accurate (for reasonably small noise levels) while the second order expansion is, and (iii) identify situations where both first and second orders are inaccurate and propose methods for improving the approximations. Chapters 2.3.1 and 2.3.2 study the performance of LASSO and other bridge regression estimators with $1 < q \leq 2$ respectively.

2.3.1 LASSO

One of the conclusions from Theorem 2.2.7 is that the first dominant term provides a good approximation of AMSE for the LASSO problem when the distribution of G does not have a large mass around 0. To test this claim we conduct the following numerical experiment. We set the parameters of our problem instances in the following way:

1. δ can take any value in $\{1.1, 1.5, 2\}$.
2. ϵ can take values in $\{0.25, 0.7\}$.
3. σ_w ranges within the interval $[0, 0.25]$.
4. the distribution of G is specified as $g(b) = 0.5\delta_1(b) + 0.5\delta_{-1}(b)$, where $\delta_a(\cdot)$ denotes a point mass at point a.

We then use the formula in Corollary 2.2.4 to calculate $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$. Finally, we compare $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$, computed numerically from (2.9) and (2.10), with its first order approximation provided in Theorem 2.2.7. The results of this experiment are summarized in Figure 2.2. As is clear in this figure, the first order expansion gives a very good approximation for AMSE over a large range of σ_w.

2.3.2 Bridge regression estimators with $1 < q \leq 2$

In this numerical experiment, we would like to vary σ_w and see under what conditions our first order or second order expansions can lead to accurate approximation of AMSE for a wide range of σ_w. Throughout this section, we set the distribution of G to $g(b) = 0.5\delta_1(b) + 0.5\delta_{-1}(b)$, as we did in Chapter 2.3.1. We then investigate different conditions by specifying various values of other parameters in our problem instances. The expansion of AMSE for $q > 1$ is presented in Theorem 2.2.5. For $q \in (1, 2)$, recall the two terms in the expansion below

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1 - 1/\delta} - \frac{\delta^{q+1}(1-\epsilon)^2(\mathbb{E}|Z|^q)^2}{(\delta-1)^{q+1}\epsilon\mathbb{E}|G|^{2q-2}}\sigma_w^{2q} + o(\sigma_w^{2q}). \tag{2.15}$$

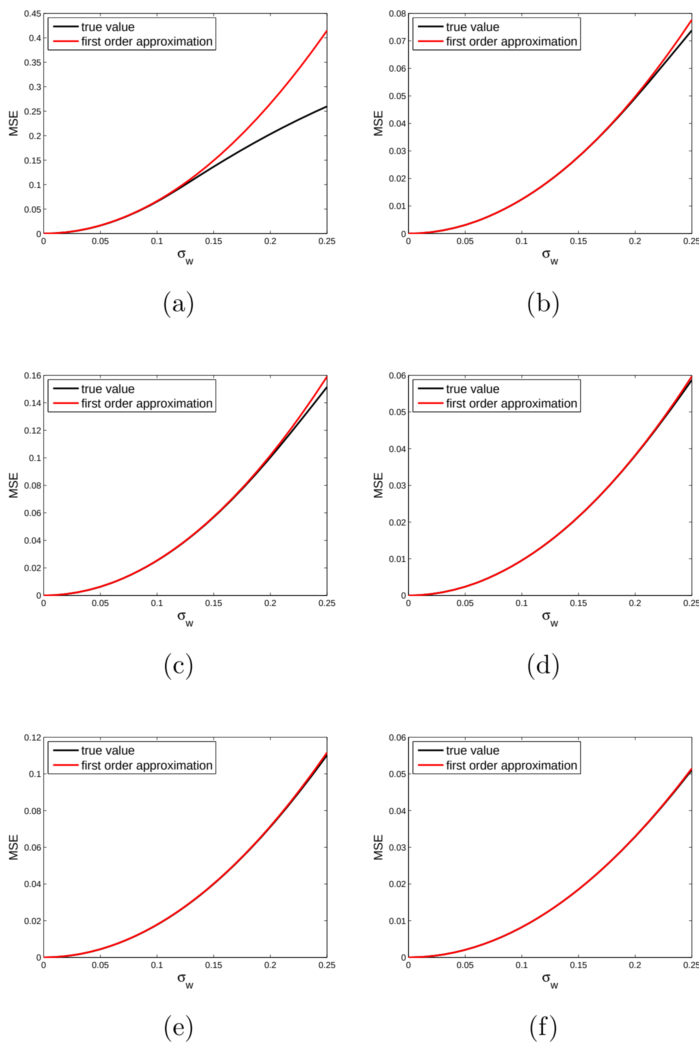
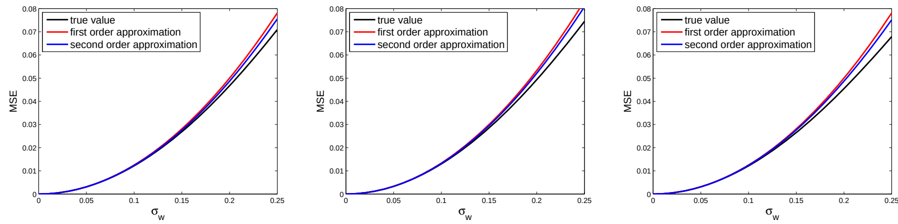


Figure 2.2: Plots of actual AMSE and its first-order approximations for (a) $\delta = 1.1$ and $\epsilon = 0.7$, (b) $\delta = 1.1$ and $\epsilon = 0.25$, (c) $\delta = 1.5$ and $\epsilon = 0.7$, (d) $\delta = 1.5$ and $\epsilon = 0.25$, (e) $\delta = 2$ and $\epsilon = 0.7$, (f) $\delta = 2$ and $\epsilon = 0.25$.



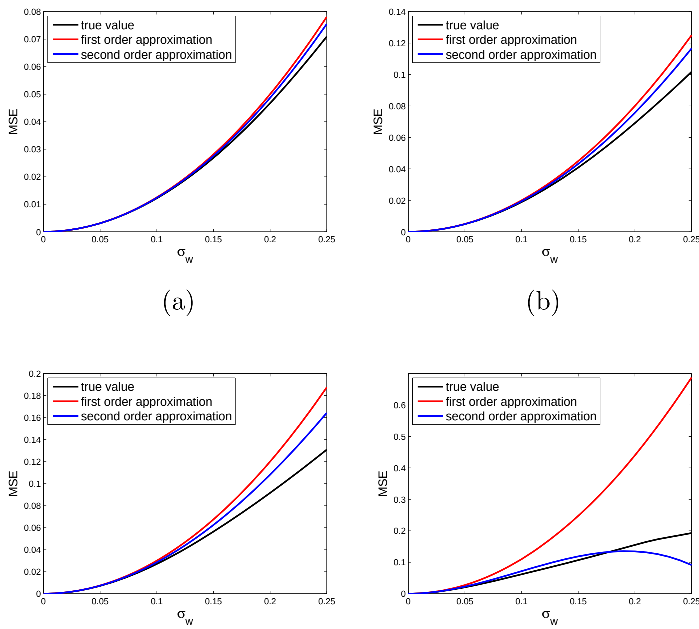
(a) (b) (c)

Figure 2.3: Plots of actual AMSE and its approximations for (a) $\delta = 5, \epsilon = 0.7, q = 1.5$, (b) $\delta = 4, \epsilon = 0.7, q = 1.6$, (c) $\delta = 5, \epsilon = 0.6, q = 1.8$.

We expect the first order term to present a good approximation over a reasonably large range of σ_w, when the second order term is sufficiently small. According to the analytical form of the second order term in (2.15), it is small if the following three conditions hold simultaneously: (i) δ is not close to 1, (ii) ϵ is not small, and (iii) q is not close to 1. Our first numerical result shown in Figure 2.3 is in agreement with this claim. In this simulation we have set three different cases for δ, ϵ and q so that they satisfy the above three conditions. The non-zero elements of β are independently drawn from $0.5\delta_1(b) + 0.5\delta_{-1}(b)$. As demonstrated in this figure, the first order term approximates AMSE accurately. Another interesting finding is that the second-order expansion provides an even better approximation.

To understand the limitation of the first order approximation, we consider the cases in which the second order term is large and suggests that at least the first order approximation is not necessarily good. This happens when either δ decreases to 1, ϵ decreases to 0 or q decreases to 1. The settings of our experiments and the results are summarized below.

1. We keep $q = 1.5$ and $\epsilon = 0.7$ fixed and study different values of $\delta \in \{5, 2, 1.5, 1.1\}$. Figure 2.4 summarizes the results of this simulation. As is clear in this figure (and is consistent with the message of the second dominant term), as we decrease δ the first order approximation becomes less accurate. The second order ap-

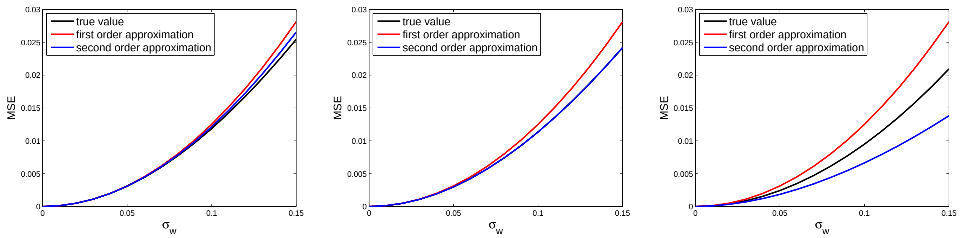


(a) (b) (c) (d)

Figure 2.4: Plots of actual AMSE and its approximations for $q = 1.5, \epsilon = 0.7$ with (a) $\delta = 5$, (b) $\delta = 2$, (c) $\delta = 1.5$ and (d) $\delta = 1.1$.

proximation in these cases is more accurate than the first order approximation. However interestingly, the second order approximation becomes less accurate as δ decreases too. These observations suggest that to have a good approximation for the values of δ that are very close to 1, although the second order approximation outperforms the first order, it may not be sufficient and higher order terms are required. Such terms can be derived with strategies similar to the ones we used in the proof of Theorem 2.2.5. Note that the insufficiency of the second order expansion partially results from the wide range of $\sigma_w \in [0, 0.25]$. If we evaluate the approximation when σ_w is small enough, we will expect the success of the second-order expansion.

2. In our second simulation, we fix $\delta = 5, \epsilon = 0.4$ and let $q \in \{1.8, 1.5, 1.1\}$. All the simulation results are summarized in Figure 2.5. As we expected, the first



(a) (b) (c)

Figure 2.5: Plots of actual AMSE and its approximations for $\delta = 5, \epsilon = 0.4$ with (a) $q = 1.8$, (b) $q = 1.5$, and (c) $q = 1.1$.

order approximation becomes less accurate when q decreases. Furthermore, we notice that when q is very close to 1 (check $q = 1.1$ in the figure), even the second order approximation is not necessarily good. This again calls for higher order approximation of the AMSE.

3. For the last simulation, we fix $\delta = 5$, $q = 1.8$, and let $\epsilon \in \{0.7, 0.5, 0.3, 0.1\}$. Our simulation results are presented in Figure 2.6. We see that as ϵ decreases the first order approximation becomes less accurate. The second order approximation is always better than the first one. Moreover, we observe that when ϵ is very close to 0 (check $\epsilon = 0.1$ in the figure), even the second order approximation is not necessarily sufficient. As we discussed in the previous two simulations, we might need higher order approximation of the AMSE in such cases.

2.3.3 Discussion

Firstly, our numerical studies confirm that the first order term gives good approximations of AMSE for LASSO in the case where the distribution of non-zero elements of β is bounded away from zero. Secondly, as the numerical results for $1 < q \leq 2$ demonstrate, while the second order approximation always improves over the first order term and works well in many cases, in the following situations it may not provide

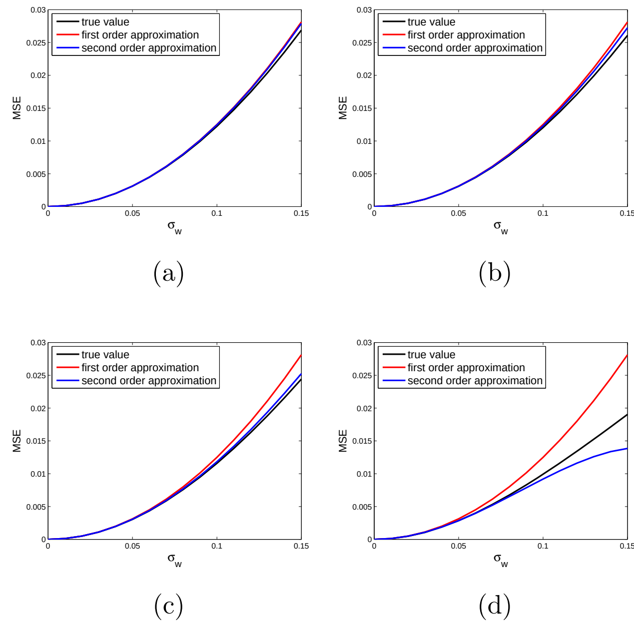


Figure 2.6: Plots of actual AMSE and its approximations for $\delta = 5$ and $q = 1.8$ with (a) $\epsilon = 0.7$, (b) $\epsilon = 0.5$, (c) $\epsilon = 0.3$, and (d) $\epsilon = 0.1$.

very accurate evaluation of AMSE: (i) when δ is close to 1, (ii) ϵ is close to zero, and (iii) q is close to 1. In such cases, the value of the second order term becomes large and hence the approximation is only accurate for very small value of σ_w. The remedy that one can propose is to derive higher order expansions. Such terms can be calculated with the same strategy that we used to obtain the second dominant term.

2.4 Related works

2.4.1 Other phase transition analyses and $n/p \to \delta$ asymptotic results

The asymptotic framework that we considered in this thesis evolved in a series of papers by Donoho and Tanner (Donoho and Tanner, 2005b; Donoho, 2004, 2006b; Donoho and Tanner, 2005a). This framework was used before on similar problems in engineering and physics (Guo and Verdú, 2005; Tanaka, 2002; Coolen, 2005). Donoho and Tanner characterized the phase transition curve for LASSO and some of its variants. Inspired by this framework, many researchers started exploring the performance of different algorithms or estimates under this asymptotic settings (STOJNIC, 2009; Amelunxen et al., 2014; Thrampoulidis et al., 2016; El Karoui et al., 2013; Karoui, 2013; Donoho and Montanari, 2013; Donoho et al., 2013; Donoho and Montanari, 2015; Bradic and Chen, 2015; Donoho et al., 2011; Zheng et al., 2017; Rangan et al., 2012; Krzakala et al., 2012; Bayati and Montanari, 2011, 2012).

Our work performs the analysis of LQLS under such asymptotic framework. Also, we adopt the message passing analysis that was developed in a series of papers (Donoho et al., 2011, 2009; Maleki, 2010; Bayati and Montanari, 2011, 2012). The notion of phase transition we consider is similar to the one introduced in Donoho et al. (2011). However, there are three major differences: (i) The analysis of Donoho et al. (2011) is performed for LASSO, while we have generalized the analysis to any LQLS with $q \in [0, \infty)$. (ii) The analysis of Donoho et al. (2011) is performed on the least favorable distribution for LASSO, while here we characterize the effect of the distribution of G on the AMSE as well. (iii) Finally, Donoho et al. (2011) is only concerned with the first dominant term in AMSE of LASSO, while we derive the second dominant term whose importance has been discussed in the last few sections.

Another line of research that has connections with our analysis for $q \geq 1$ is presented in a series of papers (Oymak et al., 2013; Oymak and Hassibi, 2016; Thrampoulidis et al., 2016). In Thrampoulidis et al. (2016) the authors have derived a minimax formulation that (if it has a unique solution and is solved) can give an accurate characterization of the asymptotic mean square error. Compared with Theorem 2.2.2 in this thesis, that result works for more general penalized M-estimators, while Theorem 2.2.2 holds for general pseudo-Lipschitz loss functions. Furthermore, Oymak et al. (2013); Oymak and Hassibi (2016) proposed a geometric approach to characterize $\lim_{\sigma_w^2 \rightarrow 0} \frac{\|\hat{\beta}-\beta\|_2^2}{\sigma_w^2}$. We can consider such result as the first-order expansion (or equivalently phase transition analysis) that we discussed in this chapter.

Several researchers have also worked on the analysis of LQLS for $q < 1$ (Kabashima et al., 2009; Rangan et al., 2012; Stojnic, 2013; Wang et al., 2011). Both Wang et al. (2011) and Stojnic (2013) performed phase transition analysis. The characterization of phase transition curve in Stojnic (2013) is only accurate for the case $q = 0$. Also, the analysis of Wang et al. (2011) is sharp only for $\delta \to 1$. Our work derives the exact value of the curve for any value of $0 \leq q < 1$ and present accurate calculations of AMSE in the presence of noise. Unlike the two papers, our analysis is based on replica method and hence is not fully rigorous yet. Replica method has been employed for studying (1.1) in Kabashima et al. (2009); Rangan et al. (2012) to derive the fixed point equations that describe the performance of $\hat{\beta}(\lambda, q)$ (under the asymptotic settings). To provide fair comparison of the performance of $\hat{\beta}(\lambda, q)$ among different $q \in [0, 1)$, one should analyze the fixed points of these equations under the optimal tuning of the parameter λ. Such analysis is missing in both papers.

2.4.2 Other analysis frameworks

One of the first papers that compared the performance of penalization techniques is Hoerl and Kennard (1970) which showed that there exists a value of λ with which Ridge regression, i.e. LQLS with $q = 2$, outperforms the vanilla least squares estimator. Since then, many more regularizers have been introduced to the literature each with a certain purpose. For instance, we can mention LASSO (Tibshirani, 1996), elastic net (Zou and Hastie, 2005), SCAD (Fan and Li, 2001), bridge regression (Frank and Friedman, 1993), and more recently SLOPE (Bogdan et al., 2015). There has been a large body of work on studying all these regularization techniques. We partition all the work into the following categories and explain what in each category has been done about the bridge regression:

(i) Simulation results: One of the main motivations for our work comes from the nice simulation study of the bridge regression presented in Fu (1998). This paper finds the optimal values of λ and q by generalized cross validation and compares the performance of the resulting estimator with both LASSO and ridge. The main conclusion is that the bridge regression can outperform both LASSO and ridge. Given our results we see that if sparsity is present in β, then smaller values of q perform better than ridge (in their second dominant term).

(ii) Asymptotic study: Knight and Fu (Knight and Fu, 2000) studied the asymptotic properties of bridge regression under the setting where $n \to \infty$, while p is fixed. They established the consistency and asymptotic normality of the estimates under quite general conditions. Huang et al. (Huang et al., 2008) studied LQLS for $q < 1$ under a high-dimensional asymptotic setting in which p grows with n but is still assumed to be less than n. They not only derived the asymptotic distribution of the estimators, but also proved LQLS has oracle properties in the sense of Fan and Li (Fan and Li, 2001). They have also considered the case $p > n$, and have shown that under partial orthogonality assumption on X, bridge regression distinguishes correctly between covariates with zero and non-zero coefficients. Note that under the asymptotic regime in this thesis, both LASSO and the other bridge estimators have false discoveries (Su et al., 2015) and possibly non-zero AMSE. Hence, they may not provide consistent estimates. Finally, the performance of LASSO under a variety of conditions has been studied extensively. We refer the reader to Bühlmann and Van De Geer (2011) for the review of those results.

(iii) Non-asymptotic bounds: One of the successful approaches that has been employed for studying the performance of regularization techniques such as LASSO is the minimax analysis (Bickel et al., 2009; Raskutti et al., 2011). We refer the reader to Bühlmann and Van De Geer (2011) for a complete list of references on this direction. In this minimax approach, a lower bound for the prediction error or mean square error of any estimation technique is first derived. Then a specific estimate, like the one returned by LASSO, is considered and an upper bound is derived assuming the design matrices satisfy certain conditions such as restrictive eigenvalue assumption (Bickel et al., 2009; Koltchinskii, 2009), restricted isometry condition (Candès, 2008), or coherence conditions (Bunea et al., 2007). These conditions can be confirmed for matrices with iid subgaussian elements. Based on these evaluations, if the order of the upper bound for the estimate under study matches the order of the lower bound, we can claim that the estimate (e.g. LASSO) is minimax rate-optimal. This approach has some advantages and disadvantages compared to our asymptotic approach: (i) It works under more general conditions. (ii) It provides information for any sample size. The price paid in the minimax analysis is that the constants derived in the results are usually not sharp and hence many schemes have similar guarantees and cannot be compared to each other. Our asymptotic framework looses the generality and in return gives sharp constants that can then be used in evaluating and comparing different schemes as we did in this chapter. Along similar directions, Koltchinskii (2009) has studied the penalized empirical risk minimization with ℓ_q penalties for the values of $q \in [1, 1 + \frac{1}{\log p}]$ and has found upper bounds on the excess risk of these estimators (oracle inequalities). Chartrand and Staneva (2008) has employed the popular analysis tool,i.e., restricted isometry property and derived a lower bound for the number of measurements required by (1.1) to recover β accurately. Similar to minimax analysis, although the results of these analyses enjoy generality, they suffer from loose constants that impede an accurate comparisons of different bridge estimators.

2.5 Proofs of the main results

2.5.1 Organization

This section contains all the proofs of the results that have not been covered in this chapter. We outline the structure of this section to help readers find the materials they are interested in. The organization is as follows:

1. Chapter 2.5.2 includes the proof of Theorem 2.2.7. Although some techniques used in the proofs of the most important results including Theorems 2.2.5, 2.2.7, 2.2.8, 2.2.10 and 2.2.11 are quite different, the roadmap remains the same. Hence we put ahead the proof of Theorem 2.2.7, the easiest one, and suggest readers to first read it. Once this relatively simple proof is clear, the other more complicated ones will be easier to read.

2. Chapter 2.5.3 covers several important properties of the proximal operator function $\eta_q(u; \chi)$. These properties will later be extensively used in the proofs.

3. Chapter 2.5.4 proves Theorem 2.2.2. This theorem characterizes the asymptotic mean square error of LQLS estimators with $q \in [1, \infty)$.

4. Chapter 2.5.5 proves Theorem 2.2.3. This theorem characterizes the asymptotic mean square error of LQLS estimators with $q \in [0, 1)$.

5. Chapter 2.5.6 includes the proof of Corollary 2.2.4. Such corollary provides us a simplified formula of asymptotic mean square error under optimal tuning.

6. Chapter 2.5.7 includes the proof of Theorem 2.2.5, one of the main results in this chapter. The theorem derives the second-order expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ for $q \in (1, \infty)$. We recommend interested readers to read the proof in Chapter 2.5.2 before Chapter 2.5.7.

7. Chapter 2.5.8 contains the proof of Theorem 2.2.6. This theorem identifies the necessary condition for successful recovery with $q \in (1, \infty)$. Phase transition is implied by this theorem together with Theorem 2.2.5.

8. Chapter 2.5.9 proves Theorem 2.2.8. The proof of this theorem is along the same lines as the proof of Theorem 2.2.7 presented in Chapter 2.5.2. We suggest the reader to study that section before studying this one.

9. Chapter 2.5.10 proves Theorem 2.2.9. The proof is essentially the same as the proof of Theorem 2.2.6. Since we do not repeat the detailed arguments, readers may want to study Chapter 2.5.8 first.

10. Chapter 2.5.11 includes the proof of Theorem 2.2.10. The theorem derives the second-order expansion of AMSE for $q \in (0, 1)$.

11. Chapter 2.5.12 includes the proof of Theorem 2.2.11. The theorem derives the second-order expansion of AMSE for $q = 0$.

12. Chapter 2.5.13 proves Theorem 2.2.12. This theorem identifies the necessary condition for successful recovery with $q \in [0, 1)$. The proof is similar to that of Theorem 2.2.6. Since we do not repeat the arguments, we refer the reader to Chapter 2.5.8.

2.5.2 Proof of Theorem 2.2.7

2.5.2.1 Roadmap of the proof

Since the proof of this result has several steps and is long, we lay out the roadmap of the proof here to help readers navigate through the details. According to Corollary 2.2.4 (let us accept Corollary 2.2.4 for the moment; its proof will be fully presented in Chapter 2.5.6), in order to evaluate $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$ as $\sigma_w \to 0$, the crucial step is to characterize $\bar{\sigma}$ from the following equation

$$\bar{\sigma}^2 = \sigma_\omega^2 + \frac{1}{\delta}\min_{\chi\geq 0}\mathbb{E}_{B,Z}[(\eta_1(B+\bar{\sigma}Z;\chi)-B)^2]. \tag{2.16}$$

To study (2.16), the key part is to analyze the term $\min_{\chi\geq 0}\mathbb{E}_{B,Z}[(\eta_1(B+\bar{\sigma}Z;\chi)-B)^2]$. A useful fact that we will prove in Chapter 2.5.2.4 can simplify the analysis of (2.16): The condition $\delta > M_1(\epsilon)$ implies that $\bar{\sigma}\to 0$, as $\sigma_w \to 0$. Hence one of the main steps of this proof is to derive the convergence rate of $\min_{\chi\geq 0}\mathbb{E}_{B,Z}[(\eta_1(B+\sigma Z;\chi)-B)^2]$, as $\sigma\to 0$. Once we obtain that rate, we then characterize the convergence rate for $\bar{\sigma}$ as $\sigma_w \to 0$ from (2.16). Finally we connect $\bar{\sigma}$ to $\text{AMSE}(\lambda_{*,1},1,\sigma_w)$ based on Corollary 2.2.4, and derive the expansion for $\text{AMSE}(\lambda_{*,1},1,\sigma_w)$ as $\sigma_w\to 0$. We introduce the following notations:

$$R(\chi,\sigma)=\mathbb{E}[(\eta_1(B/\sigma+Z;\chi)-B/\sigma)^2], \quad \chi^*(\sigma)=\arg\min_{\chi\geq 0} R(\chi,\sigma),$$

where we have suppressed the subscript B,Z in $\mathbb{E}$ for notational simplicity. According to Mousavi et al. (2017), $R(\chi,\sigma)$ is a quasi-convex function of χ and has a unique global minimizer. Hence $\chi^*(\sigma)$ is well defined. It is straightforward to confirm

$$\min_{\chi\geq 0}\mathbb{E}_{B,Z}[(\eta_1(B+\sigma Z;\chi)-B)^2]=\sigma^2 R(\chi^*(\sigma),\sigma).$$

Throughout the proof, we may write χ^* for $\chi^*(\sigma)$ when no confusion is caused, and we use $F(g)$ to denote the distribution function of $|G|$. The rest of the proof of Theorem 2.2.7 is organized in the following way:

1. We first prove $R(\chi^*(\sigma),\sigma)\to M_1(\epsilon)$, as $\sigma\to 0$ in Chapter 2.5.2.2.

2. We further bound the convergence rate of $R(\chi^*(\sigma),\sigma)$ in Chapter 2.5.2.3.

3. We finally utilize the convergence rate bound derived in Chapter 2.5.2.3 to characterize the convergence rate of $\bar{\sigma}$ and then derive the expansion for $\text{AMSE}(\lambda_{*,1},1,\sigma_w)$ in Chapter 2.5.2.4.

2.5.2.2 Proof of $R(\chi^*(\sigma), \sigma) \to M_1(\epsilon)$, as $\sigma \to 0$

Our goal in this section is to prove the following lemma.

Lemma 2.5.1. *Suppose* $\mathbb{E}|G|^2 < \infty$, *then* $\lim_{\sigma\to 0} \chi^*(\sigma) = \chi^{**}$ *and*

$$\lim_{\sigma\to 0} R(\chi^*(\sigma), \sigma) = (1-\epsilon)\mathbb{E}(\eta_1(Z;\chi^{**}))^2 + \epsilon(1+(\chi^{**})^2),$$

where $\chi = \chi^{**}$ *is the unique minimizer of* $(1-\epsilon)\mathbb{E}(\eta_1(Z;\chi))^2 + \epsilon(1+\chi^2)$ *over* $[0,\infty)$, *and* $Z \sim N(0,1)$.

Proof. By taking derivatives, it is straightforward to verify that $(1-\epsilon)\mathbb{E}(\eta_1(Z;\chi))^2 + \epsilon(1+\chi^2)$, as a function of χ over $[0,\infty)$, is strongly convex and has a unique minimizer. Hence χ^{**} is well defined.

We first claim that $\chi^*(\sigma_n)$ is bounded for any given sequence $\sigma_n \to 0$. Otherwise there exists an unbounded subsequence $\chi^*(\sigma_{n_k}) \to +\infty$ with $\sigma_{n_k} \to 0$. Since the distribution of G does not have point mass at zero and

$$\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) = \text{sign}(G/\sigma_{n_k} + Z)(|G/\sigma_{n_k} + Z| - \chi^*(\sigma_{n_k}))_+,$$

it is not hard to conclude that

$$|\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k}| \to +\infty, a.s.$$

By Fatou's lemma, we then have

$$R(\chi^*(\sigma_{n_k}), \sigma_{n_k}) \geq \epsilon\mathbb{E}(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k})^2 \to +\infty. \quad (2.17)$$

On the other hand, the optimality of $\chi^*(\sigma_{n_k})$ implies

$$R(\chi^*(\sigma_{n_k}), \sigma_{n_k}) \leq R(0, \sigma_{n_k}) = 1,$$

contradicting the unboundedness in (2.17).

We next show the sequence $\chi^*(\sigma_n)$ converges to a finite constant, for any $\sigma_n \to 0$. Taking a convergent subsequence $\chi^*(\sigma_{n_k})$, due to the boundedness of $\chi^*(\sigma_n)$, the limit of the subsequence is finite. Call it $\tilde{\chi}$. Note that

$$\begin{aligned}
&\mathbb{E}(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k})^2 \\
&= 1 + \mathbb{E}(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k} - Z)^2 + \\
&\quad 2\mathbb{E}Z(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k} - Z).
\end{aligned}$$

Since $\eta_1(u; \chi) = \text{sign}(u)(|u| - \chi)_+$, we have the following three inequalities:

$$\begin{aligned}
&|\eta_1(Z; \chi^*(\sigma_{n_k}))|^2 \leq |Z|^2, \\
&(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k} - Z)^2 \leq (\chi^*(\sigma_{n_k}))^2, \\
&|Z(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k} - Z)| \leq |Z|\chi^*(\sigma_{n_k}).
\end{aligned}$$

Furthermore, all the terms on the right hand side of the above inequalities are integrable. Therefore we can apply the Dominated Convergence Theorem (DCT) to obtain

$$\begin{aligned}
&\lim_{n_k \to \infty} R(\chi^*(\sigma_{n_k}), \sigma_{n_k}) \\
&= \lim_{n_k \to \infty} (1-\epsilon)\mathbb{E}(\eta_1(Z; \chi^*(\sigma_{n_k})))^2 + \epsilon \mathbb{E}(\eta_1(G/\sigma_{n_k} + Z; \chi^*(\sigma_{n_k})) - G/\sigma_{n_k})^2 \\
&= (1-\epsilon)\mathbb{E}(\eta_1(Z; \tilde{\chi}))^2 + \epsilon(1 + \tilde{\chi}^2).
\end{aligned}$$

Moreover, since $\chi^*(\sigma_{n_k})$ is the optimal threshold value for $R(\chi, \sigma_{n_k})$,

$$\lim_{n_k \to \infty} R(\chi^*(\sigma_{n_k}), \sigma_{n_k}) \leq \lim_{n_k \to \infty} R(\chi^{**}, \sigma_{n_k}) = (1-\epsilon)\mathbb{E}(\eta_1(Z; \chi^{**}))^2 + \epsilon(1 + (\chi^{**})^2)$$

Combining the last two limiting results, we can conclude $\tilde{\chi} = \chi^{**}$. Since $\chi^*(\sigma_{n_k})$ is an arbitrary convergent subsequence, this implies that the sequence $\chi^*(\sigma_n)$ converges to χ^{**} as well. This is true for any $\sigma_n \to 0$, hence $\chi^*(\sigma) \to \chi^{**}$, as $\sigma \to 0$. $\lim_{\sigma \to 0} R(\chi^*(\sigma), \sigma)$ can then be directly derived. □

2.5.2.3 Bounding the convergence rate of $R(\chi^*(\sigma), \sigma)$

In Chapter 2.5.2.2 we have shown $R(\chi^*(\sigma), \sigma) \to M_1(\epsilon)$ as $\sigma \to 0$. Our goal in this section is to bound the difference $R(\chi^*(\sigma), \sigma) - M_1(\epsilon)$. For that purpose, we first bound the convergence rate of $\chi^*(\sigma)$.

Lemma 2.5.2. *Suppose* $\mathbb{P}(|G| \geq \mu) = 1$ *with* μ *being a positive constant and* $\mathbb{E}|G|^2 < \infty$*, then as* $\sigma \to 0$

$$|\chi^*(\sigma) - \chi^{**}| = O(\phi(-\mu/\sigma + \chi^{**})),$$

where $\phi(\cdot)$ *is the density function of the standard normal.*

Proof. Since $\chi = \chi^*(\sigma)$ minimizes $R(\chi, \sigma)$, we have $\frac{\partial R(\chi^*(\sigma),\sigma)}{\partial \chi} = 0$, which gives the following expression for $\chi^*(\sigma)$:

$$\chi^*(\sigma) = \frac{2(1-\epsilon)\phi(\chi^*) + \epsilon\mathbb{E}\phi(\chi^* - G/\sigma) + \epsilon\mathbb{E}\phi(\chi^* + G/\sigma)}{2(1-\epsilon)\int_{\chi^*}^{\infty}\phi(z)dz + \epsilon\mathbb{E}\int_{\chi^*-G/\sigma}^{\infty}\phi(z)dz + \epsilon\mathbb{E}\int_{-\infty}^{-\chi^*-G/\sigma}\phi(z)dz}.$$

Letting σ go to zero on both sides in the above equation, we then obtain

$$\chi^{**} = \frac{2(1-\epsilon)\phi(\chi^{**})}{2(1-\epsilon)\int_{\chi^{**}}^{\infty}\phi(z)dz + \epsilon},$$

where we have applied Dominated Convergence Theorem (DCT). To bound $|\chi^*(\sigma) - \chi^{**}|$, we first bound the convergence rate of the terms in the expression of $\chi^*(\sigma)$. A direct application of the mean value theorem leads to

$$\phi(\chi^*) - \phi(\chi^{**}) = (\chi^{**} - \chi^*)\tilde{\chi}\phi(\tilde{\chi}), \tag{2.18}$$

$$\int_{\chi^*}^{\infty}\phi(z)dz - \int_{\chi^{**}}^{\infty}\phi(z)dz = (\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}), \tag{2.19}$$

with $\tilde{\chi}, \tilde{\tilde{\chi}}$ being two numbers between χ^* and χ^{**}. We now consider the other four terms. By the condition $\mathbb{P}(|G| \geq \mu) = 1$, we can conclude that for sufficiently small σ

$$\mathbb{E}\phi(\chi^* - G/\sigma) \leq \mathbb{E}\phi(\chi^* - |G|/\sigma) \leq \phi(\mu/\sigma - \chi^*), \tag{2.20}$$

$$\mathbb{E}\phi(\chi^* + G/\sigma) \leq \mathbb{E}\phi(\chi^* - |G|/\sigma) \leq \phi(\mu/\sigma - \chi^*). \tag{2.21}$$

Moreover, it is not hard to derive

$$1 - \mathbb{E}\int_{\chi^*-G/\sigma}^{\infty}\phi(z)dz - \mathbb{E}\int_{-\infty}^{-\chi^*-G/\sigma}\phi(z)dz \tag{2.22}$$
$$= \int_0^{\infty}\int_{-\chi^*-g/\sigma}^{\chi^*-g/\sigma}\phi(z)dzdF(g) \leq \int_{-\chi^*-\mu/\sigma}^{\chi^*-\mu/\sigma}\phi(z)dz \leq 2\chi^*\phi(\mu/\sigma - \chi^*),$$

where to obtain the last two inequalities we have used the condition $\mathbb{P}(|G| \geq \mu) = 1$ and the fact $\chi^* - \mu/\sigma < 0$ for σ small enough. We are now in the position to bound $|\chi^*(\sigma) - \chi^{**}|$. Define the following notations:

$$\begin{aligned} e_1 &\triangleq \epsilon\mathbb{E}\int_{\chi^*-G/\sigma}^{\infty} \phi(z)dz + \epsilon\mathbb{E}\int_{-\infty}^{-\chi^*-G/\sigma} \phi(z)dz - \epsilon, \\ e_2 &\triangleq \epsilon\mathbb{E}\phi(\chi^* - G/\sigma) + \epsilon\mathbb{E}\phi(\chi^* + G/\sigma), \\ S &\triangleq 2(1-\epsilon)\phi(\chi^{**}), \quad T \triangleq 2(1-\epsilon)\int_{\chi^{**}}^{\infty} \phi(z)dz + \epsilon. \end{aligned}$$

Using the new notations and Equations (2.18) and (2.19), we obtain

$$\chi^*(\sigma) = \frac{S + 2(1-\epsilon)(\chi^{**} - \chi^*)\tilde{\chi}\phi(\tilde{\chi}) + e_2}{T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1}, \quad \chi^{**} = \frac{S}{T}.$$

Hence we can do the following calculations:

$$\begin{aligned} \chi^*(\sigma) - \chi^{**} &= \frac{S + 2(1-\epsilon)(\chi^{**} - \chi^*)\tilde{\chi}\phi(\tilde{\chi}) + e_2}{T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1} - \frac{S}{T} \\ &= \frac{2(1-\epsilon)(\chi^{**} - \chi^*)\tilde{\chi}\phi(\tilde{\chi}) + e_2}{T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1} - \\ &\quad \frac{S(2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1)}{T(T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1)} \\ &= \frac{2(1-\epsilon)(\chi^{**} - \chi^*)(\tilde{\chi}\phi(\tilde{\chi}) - \chi^{**}\phi(\tilde{\tilde{\chi}}))}{T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1} + \\ &\quad \frac{e_2 - \chi^{**}e_1}{T + 2(1-\epsilon)(\chi^{**} - \chi^*)\phi(\tilde{\tilde{\chi}}) + e_1}. \end{aligned} \tag{2.23}$$

From (2.23) we obtain

$$\begin{aligned} &(\chi^*(\sigma) - \chi^{**})\left(1 + \frac{2(1-\epsilon)(\tilde{\chi}\phi(\tilde{\chi}) - \chi^{**}\phi(\tilde{\tilde{\chi}}))}{T + 2(1-\epsilon)(\chi^{**} - \chi^*(\sigma))\phi(\tilde{\tilde{\chi}}) + e_1}\right) \\ &= \frac{e_2 - \chi^{**}e_1}{T + 2(1-\epsilon)(\chi^{**} - \chi^*(\sigma))\phi(\tilde{\tilde{\chi}}) + e_1}. \end{aligned} \tag{2.24}$$

Note that in the above expression we have $\tilde{\chi} \to \chi^{**}$ and $\tilde{\tilde{\chi}} \to \chi^{**}$ since $\chi^*(\sigma) \to \chi^{**}$. Therefore, we conclude that $\tilde{\chi}\phi(\tilde{\chi}) - \chi^{**}\phi(\tilde{\tilde{\chi}}) \to 0$ and $(\chi^{**} - \chi^*(\sigma))\phi(\tilde{\tilde{\chi}}) \to 0$. Moreover, since (2.20), (2.21) and (2.22) together show both e_1 and e_2 go to 0 exponentially fast, we conclude from (2.24) that $(\chi^*(\sigma) - \chi^{**})/\sigma \to 0$. This enables us to proceed

$$\begin{aligned}
&\lim_{\sigma \to 0} \frac{|\chi^*(\sigma) - \chi^{**}|}{\phi(\mu/\sigma - \chi^{**})} = \lim_{\sigma \to 0} \frac{|\chi^*(\sigma) - \chi^{**}|}{\phi(\mu/\sigma - \chi^{*})} \overset{(a)}{=} \lim_{\sigma \to 0} \frac{|e_2 - \chi^{**} e_1|}{T\phi(\mu/\sigma - \chi^*)} \\
\overset{(b)}{\leq}\ &\lim_{\sigma \to 0} \frac{2\epsilon(1 + \chi^*(\sigma)\chi^{**})\phi(\mu/\sigma - \chi^*)}{T\phi(\mu/\sigma - \chi^*)} = \frac{2\epsilon(1 + (\chi^{**})^2)}{T}.
\end{aligned}$$

We have used (2.24) to obtain (a). We derived (b) by the following steps:

1. According to (2.22), $|e_1| \leq 2\epsilon\chi^*\phi(\mu/\sigma - \chi^*)$.

2. According to (2.20) and (2.21), $|e_2| \leq 2\epsilon\phi(\mu/\sigma - \chi^*)$.

This completes the proof of Lemma 2.5.2. □

The next step is to bound the convergence rate of $R(\chi^*(\sigma), \sigma)$ based on the convergence rate of $\chi^*(\sigma)$ we have derived in Lemma 2.5.2.

Lemma 2.5.3. *Suppose* $\mathbb{P}(|G| \geq \mu) = 1$ *with* μ *being a positive constant and* $\mathbb{E}|G|^2 < \infty$*, then as* $\sigma \to 0$

$$|R(\chi^*(\sigma), \sigma) - M_1(\epsilon)| = O(\phi(\mu/\sigma - \chi^{**})),$$

where $\phi(\cdot)$ *is the density function of the standard normal.*

Proof. We recall the two quantities:

$$\begin{aligned}
M_1(\epsilon) &= (1-\epsilon)\mathbb{E}(\eta_1(Z; \chi^{**}))^2 + \epsilon(1 + (\chi^{**})^2), \qquad (2.25)\\
R(\chi^*(\sigma), \sigma) &= (1-\epsilon)\mathbb{E}(\eta_1(Z; \chi^{*}))^2 + \\
&\quad \epsilon[1 + \mathbb{E}(\eta_1(G/\sigma + Z; \chi^*) - G/\sigma - Z)^2] \\
&\quad + 2\epsilon\mathbb{E}Z(\eta_1(G/\sigma + Z; \chi^*) - G/\sigma - Z). \qquad (2.26)
\end{aligned}$$

We bound $|R(\chi^*(\sigma), \sigma) - M_1(\epsilon)|$ by bounding the difference between the corresponding terms in (2.26) and (2.25). From the proof of Lemma 2.5.2 we know $e_1 < 0$ and $e_2 > 0$.

Hence (2.24) implies $\chi^*(\sigma) > \chi^{**}$ for small enough σ. We start with

$$\begin{aligned}
& |\mathbb{E}(\eta_1(Z;\chi^*))^2 - \mathbb{E}(\eta_1(Z;\chi^{**}))^2| \qquad (2.27)\\
= \ & |\mathbb{E}(\eta_1(Z;\chi^*) - \eta_1(Z;\chi^{**}))(\eta_1(Z;\chi^*) + \eta_1(Z;\chi^{**}))| \\
\leq \ & \mathbb{E}[|\chi^* - \chi^{**} + \chi^*\mathbb{I}(|Z| \in (\chi^{**},\chi^*))| \cdot |\eta_1(Z;\chi^*) + \eta_1(Z;\chi^{**})|] \\
\overset{(a)}{\leq} \ & 2(\chi^* - \chi^{**}) \cdot \mathbb{E}|Z| + 2\chi^*\mathbb{E}[\mathbb{I}(|Z| \in (\chi^{**},\chi^*))|Z|] \\
\leq \ & 2(\chi^* - \chi^{**}) \cdot \mathbb{E}|Z| + 4\chi^*(\chi^* - \chi^{**})\tilde{\chi}\phi(\tilde{\chi}) = O(\phi(\mu/\sigma - \chi^{**})),
\end{aligned}$$

where we have used the fact $|\eta_1(u;\chi)| \leq |u|$ to obtain (a); $\tilde{\chi}$ is a number between $\chi^*(\sigma)$ and χ^{**}; and the last equality is due to Lemma 2.5.2. We next bound the difference between $\mathbb{E}(\eta_1(G/\sigma + Z;\chi^*) - G/\sigma - Z)^2$ and $(\chi^{**})^2$:

$$\begin{aligned}
& |(\chi^{**})^2 - \mathbb{E}(\eta_1(G/\sigma + Z;\chi^*) - G/\sigma - Z)^2| \qquad (2.28)\\
\leq \ & |(\chi^*)^2 - \mathbb{E}(\eta_1(G/\sigma + Z;\chi^*) - G/\sigma - Z)^2| + |(\chi^{**})^2 - (\chi^*)^2|.
\end{aligned}$$

To bound the two terms on the right hand side of (2.28), first note that

$$\begin{aligned}
0 \ \leq \ & (\chi^*)^2 - \mathbb{E}(\eta_1(G/\sigma + Z;\chi^*) - G/\sigma - Z)^2 \\
= \ & \mathbb{E}[\mathbb{I}(|G/\sigma + Z| \leq \chi^*) \cdot ((\chi^*)^2 - (G/\sigma + Z)^2)] \\
\leq \ & (\chi^*)^2 \int_0^\infty \int_{-g/\sigma-\chi^*}^{-g/\sigma+\chi^*} \phi(z)dzdF(g) \\
\overset{(b)}{\leq} \ & (\chi^*)^2 \int_{-\mu/\sigma-\chi^*}^{-\mu/\sigma+\chi^*} \phi(z)dz \leq 2(\chi^*)^3\phi(\mu/\sigma - \chi^*) \\
= \ & O(\phi(\mu/\sigma - \chi^{**})), \qquad (2.29)
\end{aligned}$$

where (b) is due to the condition $\mathbb{P}(|G| \geq \mu) = 1$, and the last equality holds since $(\chi^* - \chi^{**})/\sigma \to 0$ implied by Lemma 2.5.2. Furthermore, Lemma 2.5.2 yields

$$(\chi^*)^2 - (\chi^{**})^2 = O(\phi(\mu/\sigma - \chi^{**})). \qquad (2.30)$$

Combining (2.28), (2.29), and (2.30), we obtain

$$|(\chi^{**})^2 - \mathbb{E}(\eta_1(G/\sigma + Z;\chi^*) - G/\sigma - Z)^2| = O(\phi(\mu/\sigma - \chi^{**})). \qquad (2.31)$$

Regarding the remaining term in $R(\chi^*(\sigma), \sigma)$, we can derive

$$\begin{aligned} 0 &\leq \mathbb{E}Z(G/\sigma + Z - \eta_1(G/\sigma + Z; \chi^*)) \\ &\overset{(c)}{=} \mathbb{E}(1 - \partial_1 \eta_1(G/\sigma + Z; \chi^*)) \\ &= \mathbb{P}(|G/\sigma + Z| \leq \chi^*) \overset{(d)}{=} O(\phi(\mu/\sigma - \chi^{**})). \end{aligned} \quad (2.32)$$

We have employed Stein's lemma to obtain (c). Equality (d) holds due to (2.22). Putting the results (2.27), (2.31), and (2.32) together finishes the proof. □

2.5.2.4 Deriving the expansion of AMSE$(\lambda_{*,1}, 1, \sigma_w)$

In this section we utilize the convergence rate result of $R(\chi^*(\sigma), \sigma)$ from Chapter 2.5.2.3 to derive the expansion of AMSE$(\lambda_{*,1}, 1, \sigma_w)$ in (2.13), and thus finish the proof of Theorem 2.2.7. Towards that goal, we first prove a useful lemma.

Lemma 2.5.4. *Let $\bar{\sigma}$ be the solution to the following equation:*

$$\bar{\sigma}^2 = \sigma_\omega^2 + \frac{1}{\delta} \min_{\chi \geq 0} \mathbb{E}_{B,Z}[(\eta_1(B + \bar{\sigma} Z; \chi) - B)^2]. \quad (2.33)$$

Suppose $\delta > M_1(\epsilon)$, then

$$\lim_{\sigma_w \to 0} \frac{\sigma_w^2}{\bar{\sigma}^2} = \frac{\delta - M_1(\epsilon)}{\delta}.$$

Proof. We first claim that $\mathbb{E}(\eta_1(\alpha + Z; \chi) - \alpha)^2$ is an increasing function of α, because

$$\frac{d}{d\alpha} \mathbb{E}(\eta_1(\alpha + Z; \chi) - \alpha)^2 = 2\mathbb{E}(\alpha \mathbb{I}(|\alpha + Z| \leq \chi)) \geq 0.$$

Hence we obtain

$$\mathbb{E}(\eta_1(\alpha + Z; \chi) - \alpha)^2 \leq \lim_{\alpha \to \infty} \mathbb{E}(\eta_1(\alpha + Z; \chi) - \alpha)^2 = 1 + \chi^2. \quad (2.34)$$

Inequality (2.34) then yields

$$\begin{aligned} R(\chi, \bar{\sigma}) &= (1 - \epsilon)\mathbb{E}(\eta_1(Z; \chi))^2 + \epsilon \mathbb{E}(\eta_1(G/\bar{\sigma} + Z; \chi) - G/\bar{\sigma})^2 \\ &\leq (1 - \epsilon)\mathbb{E}(\eta_1(Z; \chi))^2 + \epsilon(1 + \chi^2). \end{aligned}$$

Taking minimum over χ on both sides above gives us

$$R(\chi^*(\bar{\sigma}), \bar{\sigma}) \leq M_1(\epsilon). \tag{2.35}$$

Moreover, since $\bar{\sigma}$ is the solution of (2.33), it satisfies

$$\bar{\sigma}^2 = \sigma_w^2 + \frac{\bar{\sigma}^2}{\delta} R(\chi^*(\bar{\sigma}), \bar{\sigma}). \tag{2.36}$$

Combining (2.35) and (2.36) with the condition $\delta > M_1(\epsilon)$, we have

$$\bar{\sigma}^2 \leq \frac{\sigma_w^2}{1 - M_1(\epsilon)/\delta},$$

which leads to $\bar{\sigma} \to 0$, as $\sigma_w \to 0$. Then applying Lemma 2.5.1 shows

$$\lim_{\sigma_w \to 0} R(\chi^*(\bar{\sigma}), \bar{\sigma}) = \lim_{\bar{\sigma} \to 0} R(\chi^*(\bar{\sigma}), \bar{\sigma}) = M_1(\epsilon).$$

Diving both sides of (2.36) by $\bar{\sigma}^2$ and letting $\sigma_w \to 0$ finishes the proof. □

To complete the proof of Theorem 2.2.7, first note that Corollary 2.2.4 tells us

$$\text{AMSE}(\lambda_{*,1}, 1, \sigma_w) = \bar{\sigma}^2 R(\chi^*(\bar{\sigma}), \bar{\sigma}), \quad \sigma_w^2 = \bar{\sigma}^2 - \frac{\bar{\sigma}^2}{\delta} R(\chi^*(\bar{\sigma}), \bar{\sigma}).$$

We then have

$$\begin{aligned} & \text{AMSE}(\lambda_{*,1}, 1, \sigma_w) - \frac{\delta M_1(\epsilon)}{\delta - M_1(\epsilon)} \sigma_w^2 \\ = \;& \bar{\sigma}^2 R(\chi^*(\bar{\sigma}), \bar{\sigma}) - \frac{\delta M_1(\epsilon)}{\delta - M_1(\epsilon)} \cdot \left[\bar{\sigma}^2 - \frac{\bar{\sigma}^2}{\delta} R(\chi^*(\bar{\sigma}), \bar{\sigma}) \right] \\ = \;& \frac{\delta (R(\chi^*(\bar{\sigma}), \bar{\sigma}) - M_1(\epsilon))}{\delta - M_1(\epsilon)} \bar{\sigma}^2 \overset{(a)}{=} O(\bar{\sigma}^2 \phi(\mu/\bar{\sigma} - \chi^{**})), \end{aligned} \tag{2.37}$$

where (a) is due to Lemma 2.5.3. Finally, since $\lim_{\sigma_w \to 0} \frac{\sigma_w^2}{\bar{\sigma}^2} = \frac{\delta - M_1(\epsilon)}{\delta}$ according to Lemma 2.5.4, it is not hard to see

$$O(\bar{\sigma}^2 \phi(\mu/\bar{\sigma} - \chi^{**})) = o(\phi(\bar{\mu}/\bar{\sigma})) = o\Big(\phi\Big(\sqrt{\frac{\delta - M_1(\epsilon)}{\delta}} \frac{\tilde{\mu}}{\sigma_w}\Big)\Big), \tag{2.38}$$

where $\bar{\mu}$ and $\tilde{\mu}$ are any constants satisfying $0 \leq \tilde{\mu} < \bar{\mu} < \mu$. Results (2.37) and (2.38) together close the proof of Theorem 2.2.7.

Remark: (2.35) and (2.37) together imply that the second dominant term of $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$ is in fact negative.

2.5.3 Preliminaries on $\eta_q(u;\chi)$

This section is devoted to the properties of $\eta_q(u;\chi)$ defined as

$$\eta_q(u;\chi) \triangleq \arg\min_z \frac{1}{2}(u-z)^2 + \chi|z|^q. \tag{2.39}$$

We start with some basic properties of these functions. Since the explicit forms of $\eta_q(u;\chi)$ for $q=1$ and 2 are known: $\eta_1(u;\chi) = (|u|-\chi)\mathrm{sign(u)}\mathbb{I}(|\mathrm{u}|>\chi), \eta_2(\mathrm{u};\chi) = \frac{\mathrm{u}}{1+2\chi}$, we first focus our study on the case $1<q<2$.

Lemma 2.5.5. $\eta_q(u;\chi)$ *satisfies the following properties:*

(i) $u - \eta_q(u;\chi) = \chi q \mathrm{sign}(u)|\eta_q(u;\chi)|^{q-1}$.

(ii) $|\eta_q(u;\chi)| \le |u|$.

(iii) $\lim_{\chi\to 0}\eta_q(u;\chi) = u$ *and* $\lim_{\chi\to\infty}\eta_q(u;\chi) = 0$.

(iv) $\eta_q(-u;\chi) = -\eta_q(u;\chi)$.

(v) For $\alpha > 0$, *we have* $\eta_q(\alpha u;\alpha^{2-q}\chi) = \alpha\eta_q(u;\chi)$.

(vi) $|\eta_q(u;\chi) - \eta_q(\tilde{u},\chi)| \le |u-\tilde{u}|$.

Proof. To prove (i), we should take the derivative of $\frac{1}{2}(u-z)^2 + \chi|z|^q$ and set it to zero. Proofs of parts (ii), (iii) and (iv) are straightforward and are hence skipped. To prove (v), note that

$$\begin{aligned}\eta_q(\alpha u;\alpha^{2-q}\chi) &= \arg\min_z \frac{1}{2}(\alpha u - z)^2 + \chi\alpha^{2-q}|z|^q \\ &= \arg\min_z \frac{\alpha^2}{2}(u - z/\alpha)^2 + \chi\alpha^2|z/\alpha|^q \\ &= \alpha \arg\min_{\tilde{z}} \frac{1}{2}(u-\tilde{z})^2 + \chi|\tilde{z}|^q = \alpha\eta_q(u;\chi).\end{aligned} \tag{2.40}$$

(vi) is a standard property of proximal operators of convex functions (Parikh and Boyd, 2014). □

In many proofs, we will be dealing with derivatives of $\eta_q(u;\chi)$. To simplify the notations, we may use $\partial_1\eta_q(u;\chi), \partial_1^2\eta_q(u;\chi), \partial_2\eta_q(u;\chi), \partial_2^2\eta_q(u;\chi)$ to represent $\frac{\partial\eta_q(u,\chi)}{\partial u}, \frac{\partial^2\eta_q(u,\chi)}{\partial u^2}, \frac{\partial\eta_q(u,\chi)}{\partial \chi}, \frac{\partial^2\eta_q(u,\chi)}{\partial \chi^2}$, respectively. Our next two lemmas are concerned with differentiability of $\eta_q(u;\chi)$ and its derivatives.

Lemma 2.5.6. *For every $1 < q < 2$, $\eta_q(u;\chi)$ is a differentiable function of (u,χ) for $u \in \mathbb{R}$ and $\chi > 0$ with continuous partial derivatives. Moreover, $\partial_2\eta_q(u;\chi)$ is differentiable with respect to u, for any given $\chi > 0$.*

Proof. We start with the case $u_0, \chi_0 > 0$. The goal is to prove that $\eta_q(u;\chi)$ is differentiable at (u_0,χ_0). Since $u_0 > 0$, $\eta_q(u_0;\chi_0)$ will be positive. Then Lemma 2.5.5 part (i) shows $\eta_q(u_0;\chi_0)$ must satisfy

$$\eta_q(u_0;\chi_0) + \chi_0 q\eta_q^{q-1}(u_0;\chi_0) = u_0. \tag{2.41}$$

Define the function $F(u,\chi,v) = u - v - \chi q v^{q-1}$. Equation (2.41) says $F(u,\chi,v)$ is equal to zero at $(u_0,\chi_0,\eta_q(u_0;\chi_0))$. It is straightforward to confirm that the derivative of $F(u,\chi,v)$ with respect to v is nonzero at $(u_0,\chi_0,\eta_q(u_0;\chi_0))$. By implicit function theorem, we can conclude $\eta_q(u;\chi)$ is differentiable at (u_0,χ_0). Lemma 2.5.5 part (iv) implies that the same result holds when $u_0 < 0$. We now focus on the point $(0,\chi_0)$. Since $\eta_q(0,\chi_0) = 0$, we obtain

$$\partial_1\eta_q(0;\chi_0) = \lim_{u\to 0}\frac{|\eta_q(u;\chi_0)|}{|u|} \le \lim_{u\to 0}\frac{|u|^{1/(q-1)}}{(\chi_0 q)^{1/(q-1)}|u|} = 0.$$

where the last inequality comes from (2.41). It is straightforward to see that the partial derivative of $\eta_q(u;\chi)$ with respect to χ at $(0,\chi_0)$ exists and is equal to zero as well. So far we have proved that $\eta_q(u,\chi)$ has partial derivatives with respect to both u and χ for every $u \in \mathbb{R}, \chi > 0$. We next show the partial derivatives are continuous. For $u \neq 0$, the result comes directly from the implicit function theorem, because $F(u,\chi,v)$ is a smooth function when $v \neq 0$. We now turn to the proof when $u = 0$. By taking derivative with respect to u on both sides of (2.41), we obtain

$$\partial_1\eta_q(u;\chi) + \chi q(q-1)\eta_q^{q-2}(u;\chi)\partial_1\eta_q(u;\chi) = 1, \tag{2.42}$$

for any $u, \chi > 0$. Moreover, it is clear from (2.41) that $\eta_q(u;\chi) \to 0$, as $(u,\chi) \to (0^+, \chi_0)$. This fact combined with (2.42) yields

$$\lim_{(u,\chi)\to(0^+,\chi_0)} \partial_1 \eta_q(u;\chi) = \lim_{(u,\chi)\to(0^+,\chi_0)} \frac{1}{1+\chi q(q-1)\eta_q^{q-2}(u;\chi)} = 0.$$

Since $\partial_1\eta_q(u;\chi) = \partial_1\eta_q(-u;\chi)$ implied by Lemma 2.5.5 part (iv), we conclude

$$\lim_{(u,\chi)\to(0,\chi_0)} \partial_1 \eta_q(u;\chi) = 0.$$

The same approach can prove that the partial derivative $\partial_2\eta_q(u;\chi)$ is continuous at $(0,\chi_0)$. For simplicity we do not repeat the arguments.

We now prove the second part of the lemma. Because $F(u,\chi,v)$ is infinitely many times differentiable in any open set with $v \neq 0$, implicit function theorem further implies $\partial_2\eta_q(u;\chi)$ is differentiable at any $u \neq 0$. The rest of the proof is to show its differentiability at $u = 0$. This follows by noting $\partial_2\eta_q(0;\chi) = 0$, and

$$\begin{aligned}
\lim_{u\to 0} \frac{\partial_2\eta_q(u;\chi)}{u} &\overset{(a)}{=} \lim_{u\to 0} \frac{-q|\eta_q(u;\chi)|^{q-1}}{|u|(1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2})} \\
&= \lim_{u\to 0} \frac{-(|u| - |\eta_q(u;\chi)|)}{\chi|u|(1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2})} = 0,
\end{aligned}$$

where (a) is by taking derivative with respect to χ on both sides of (2.41), and the last two equalities above are due to Lemma 2.5.5 part (i) and (iii). □

Lemma 2.5.7. *Consider a given $\chi > 0$, then for every $1 < q < 3/2$, $\partial_1\eta_q(u;\chi)$ is a differentiable function of u for $u \in \mathbb{R}$ with continuous derivative; for $q = 3/2$, it is a weakly differentiable function of u; for $3/2 < q < 2$, $\partial_1\eta_q(u;\chi)$ is differentiable at $u \neq 0$, but is not differentiable at zero.*

Proof. As is clear from the proof of Lemma 2.5.6, the implicit function theorem guarantees that $\partial_1\eta_q(u;\chi)$ is differentiable at $u \neq 0$ with continuous derivative for $1 < q < 2$. Hence we will be focused on $u = 0$. In the proof of Lemma 2.5.6, we have derived

$$\partial_1\eta_q(u;\chi) = \frac{1}{1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2}}, \quad \text{for } u \neq 0, \tag{2.43}$$

and $\partial_1 \eta_q(0;\chi) = 0$. We thus know

$$\partial_1^2 \eta_q(0;\chi) = \lim_{u \to 0} \frac{1}{u + \chi q(q-1)u|\eta_q(u;\chi)|^{q-2}}. \tag{2.44}$$

Moreover, Lemma 2.5.5 part (i) implies

$$\lim_{u \to 0} \frac{u}{\chi q |\eta_q(u;\chi)|^{q-1}\text{sign}(u)} = 1 + \lim_{u \to 0} \frac{|\eta_q(u;\chi)|^{2-q}}{\chi q} = 1. \tag{2.45}$$

For $1 < q < 3/2$, (2.44) and (2.45) together give us

$$\partial_1^2 \eta_q(0;\chi) = \lim_{u \to 0} \frac{1}{u + \chi q(q-1)u(|u|/(\chi q))^{\frac{q-2}{q-1}}} = 0.$$

We can also calculate the limit of $\partial_1^2 \eta_q(u;\chi)$ (this second derivative can be obtained from (2.43)) as follows.

$$\lim_{u \to 0} \partial_1^2 \eta_q(u;\chi) = \lim_{u \to 0} \frac{-\chi q(q-1)(q-2)|\eta_q(u;\chi)|^{q-3}\text{sign}(u)}{(1 + \chi q(q-1)|\eta_q(u;\chi)|^{q-2})^3} = 0.$$

Therefore, $\partial_1 \eta_q(u;\chi)$ is continuously differentiable on $(-\infty, +\infty)$ for $1 < q < 3/2$. Regarding $3/2 < q < 2$, Similar calculations yield

$$\lim_{u \to 0^+} \partial_1^2 \eta_q(0;\chi) = +\infty, \quad \lim_{u \to 0^-} \partial_1^2 \eta_q(0;\chi) = -\infty.$$

Finally to prove the weak differentiability for $q = 3/2$, we show $\partial_1 \eta_q(u;\chi)$ is a Lipschitz continuous function on $(-\infty, +\infty)$. Note that for $u \neq 0$,

$$|\partial_1^2 \eta_q(u;\chi)| = \frac{\chi q(q-1)(2-q)|\eta_q(u;\chi)|^{q-3}}{(1 + \chi q(q-1)|\eta_q(u;\chi)|^{q-2})^3} \leq \frac{8}{9\chi^2},$$

and $\partial_1^2 \eta_q(0^+;\chi) = -\partial_1^2(0^-;\chi) = \frac{8}{9\chi^2}$. Mean value theorem leads to

$$|\partial_1 \eta_q(u;\chi) - \partial_1 \eta_q(\tilde{u};\chi)| \leq \frac{8}{9\chi^2}|u - \tilde{u}|, \quad \text{for } u\tilde{u} \geq 0.$$

When $u\tilde{u} < 0$, we can have

$$\begin{aligned} & |\partial_1 \eta_q(u;\chi) - \partial_1 \eta_q(\tilde{u};\chi)| = |\partial_1 \eta_q(u;\chi) - \partial_1 \eta_q(-\tilde{u};\chi)| \\ \leq \ & \frac{8}{9\chi^2}|u + \tilde{u}| \leq \frac{8}{9\chi^2}|u - \tilde{u}|. \end{aligned}$$

This completes the proof of the lemma. □

The next lemma presents some additional properties regarding the derivatives of $\eta_q(u;\chi)$.

Lemma 2.5.8. *For $1 < q < 2$, the derivatives of $\eta_q(u;\chi)$ satisfy the following properties:*

(i) $\partial_1 \eta_q(u;\chi) = \frac{1}{1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2}}$.

(ii) $\partial_2 \eta_q(u;\chi) = \frac{-q|\eta_q(u;\chi)|^{q-1}\text{sign}(u)}{1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2}}$.

(iii) $0 \leq \partial_1 \eta_q(u;\chi) \leq 1$.

(iv) *For $u > 0$, $\partial_1^2 \eta_q(u;\chi) > 0$.*

(v) $|\eta_q(u;\chi)|$ *is a decreasing function of χ.*

(vi) $\lim_{\chi\to\infty} \partial_1 \eta_q(u;\chi) = 0$.

Proof. Parts (i) (ii) have been derived in the proof of Lemma 2.5.6. Part (iii) is a simple conclusion of part (i). Part (iv) is clear from the proof of Lemma 2.5.7. Part (v) is a simple application of part (ii). Finally, part (vi) is an application of part(i) of Lemma 2.5.8 and part (iii) of Lemma 2.5.5. □

We next study $\eta_q(u;\chi)$ for the case $0 \leq q < 1$.

Lemma 2.5.9. *Denote $c_q = [2(1-q)]^{\frac{1}{2-q}} + q[2(1-q)]^{\frac{q-1}{2-q}}$. Then for $0 \leq q < 1$, $\eta_q(u;\chi) = 0$ if $0 \leq u < c_q\chi^{\frac{1}{2-q}}$, and the following holds when $u > c_q\chi^{\frac{1}{2-q}}$.*

(i) $\frac{\partial \eta_q(u;\chi)}{\partial \chi} = \frac{-q\eta_q^{q-1}(u;\chi)}{1+\chi q(q-1)\eta_q^{q-2}(u;\chi)}$.

(ii) $\frac{\partial \eta_q(u;\chi)}{\partial u} = \frac{1}{1+\chi q(q-1)\eta_q^{q-2}(u;\chi)}$.

(iii) $u - \eta_q(u;\lambda) = q\chi\eta_q^{q-1}(u;\chi)$.

(iv) $\alpha\eta_q(u;\chi) = \eta_q(\alpha u;\alpha^{2-q}\chi)$ *for $\alpha > 0$.*

Lemma 2.5.9 implies that $\eta_q(u;\chi)$ has a jump at $u = c_q\chi^{\frac{1}{2-q}}$. We define that value:

$$\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi) = \lim_{u\searrow c_q\chi^{\frac{1}{2-q}}} \eta_q(u;\chi).$$

Lemma 2.5.10. *For $0 < q < 1$, If $|\eta_q(u;\chi)| > 0$, then $|\eta_q(u;\chi)| \geq [2(1-q)]^{\frac{1}{2-q}}\chi^{\frac{1}{2-q}}$.*

Proof. We refer to Zheng et al. (2017) for the complete proof of Lemmas 2.5.9 and 2.5.10. □

2.5.4 Proof of Theorem 2.2.2

2.5.4.1 Roadmap of the proof

This section contains the proof of Theorem 2.2.2. The proof for LASSO ($q = 1$) has been shown in Bayati and Montanari (2012). We aim to extend the results to $q > 1$. We will follow similar proof strategy as the one proposed in Bayati and Montanari (2012). However, as will be described later some of the steps are more challenging for $q > 1$ (and some are easier). We only present the proof for $1 < q \leq 2$. Similar arguments hold for $q > 2$. Motivated by Bayati and Montanari (2012) we construct an approximate message passing (AMP) algorithm for solving LQLS. We then establish an asymptotic equivalence between the output of AMP and the bridge regression estimates. We finally utilize the existing asymptotic results from AMP framework to prove Theorem 2.2.2. The rest of the material is organized as follows. In Chapter 2.5.4.2, we first prove the existence and uniqueness of the solution pair to (2.5) and (2.6). In Chapter 2.5.4.3, we briefly review approximate message passing algorithms and state some relevant results that will be used later in our proof. Chapter 2.5.4.4 collects two useful results to be applied in the later proof. We describe the main proof steps in Chapter 2.5.4.5.

2.5.4.2 Solution of the fixed point equations

Lemma 2.5.11. *For any positive values of $\lambda, \delta, \sigma_w > 0$, any random variable B with finite second moment, and any $q \in [1, 2]$, there exists a unique pair $(\bar{\sigma}, \bar{\chi})$ that satisfies both (2.5) and (2.6).*

To prove the above result we we pursue the following two main steps:

1. We first show the existence of the solution. In order to do that, we first study the solution of Equation (2.5), and demonstrate that for any $\chi \in (\chi_{\min}, \infty)$ ($\chi_{\min}$ is a constant we will clarify later), there exists a unique σ_χ such that (σ_χ, χ) satisfies (2.5). We then show that by varying χ over $(\chi_{\min}, \infty)$, the range of the value of the following term

$$\chi {\sigma_\chi}^{2-q}\big(1 - \frac{1}{\delta}\mathbb{E}_{B,Z}[\eta_q'(B + \sigma_\chi Z; \chi\sigma_\chi^{2-q})]\big)$$

   covers the number λ from Equation (2.6). That means Equations (2.5) and (2.6) share at least one common solution pair (σ_χ, χ).

2. We then prove the uniqueness of the solution. The key idea is to apply Theorem 2.2.2 to evaluate the asymptotic loss of the LQLS estimates under two different pseudo-Lipschitz functions. These two quantities determine the uniqueness of both σ_χ and χ in the common solution pair (σ_χ, χ). Note that we have denoted this unique pair by $(\bar{\sigma}, \bar{\chi})$.

Before we start the details of the proof, we present Stein's lemma (Stein, 1981) that will be used several times in the proof.

Lemma 2.5.12. *Let $g : \mathbb{R} \to \mathbb{R}$ denote a weakly differentiable function. If $Z \sim N(0, 1)$ and $\mathbb{E}|g'(Z)| < \infty$, we have*

$$\mathbb{E}(Zg(Z)) = \mathbb{E}(g'(Z)),$$

where g' denotes the weak-derivative of g.

We first define a function that is closely related to Equation (2.5):

$$R_q(\chi, \sigma) \triangleq \mathbb{E}_{B,Z}[\eta_q(B/\sigma + Z; \chi) - B/\sigma]^2. \tag{2.46}$$

Note that we have used the same definition for LASSO in Chapter 2.5.2. Here we adopt a general notation $R_q(\chi, \sigma)$ to represent the function defined above for any $q \in [1, 2]$.

Lemma 2.5.13. *For $1 \leq q \leq 2$, $R_q(\chi, \sigma)$ is a decreasing function of $\sigma > 0$.*

Proof. We consider four different cases: (i) $q = 1$, (ii) $q = 2$, (iii) $1 < q \leq 3/2$, (iv) $3/2 < q < 2$.

(i) $q = 1$: Since $R_1(\chi, \sigma)$ is a differentiable function of σ, we will prove this case by showing that $\frac{\partial R_1(\chi,\sigma)}{\partial \sigma} < 0$. We have

$$\begin{aligned} \frac{\partial R_1(\chi, \sigma)}{\partial \sigma} &= -\frac{2}{\sigma^2}\mathbb{E}\left[B(\mathbb{I}(|B/\sigma + Z| > \chi) - 1)(\eta_1(B/\sigma + Z; \chi) - B/\sigma)\right] \\ &= -\frac{2}{\sigma^2}\mathbb{E}\left[\mathbb{I}(|B/\sigma + Z| \leq \chi)(B^2/\sigma)\right] < 0. \end{aligned}$$

The first equality above is due to Dominated Convergence Theorem.

(ii) $q = 2$: Since $\eta_2(u; \chi) = \frac{u}{1+2\chi}$, we have

$$\frac{\partial R_2(\chi, \sigma)}{\partial \sigma} = -\frac{8\chi^2 \mathbb{E}|B|^2}{(1 + 2\chi)^2 \sigma^3} < 0.$$

(iii) $1 < q \leq 3/2$: The strategy for this case is similar to that of the last two cases. We show that the derivative $\frac{\partial R_q(\chi,\sigma)}{\partial \sigma} < 0$.

$$\begin{aligned} \frac{\partial R_q(\chi, \sigma)}{\partial \sigma} &\overset{(a)}{=} 2\mathbb{E}\left[(\eta_q(B/\sigma + Z; \chi) - B/\sigma)(\partial_1 \eta_q(B/\sigma + Z; \chi) - 1)(-B/\sigma^2)\right] \\ &= 2\mathbb{E}\left[(\eta_q(B/\sigma + Z; \chi) - B/\sigma - Z)(\partial_1 \eta_q(B/\sigma + Z; \chi) - 1)(-B/\sigma^2)\right] \\ &\quad + 2\mathbb{E}\left[Z(\partial_1 \eta_q(B/\sigma + Z; \chi) - 1)(-B/\sigma^2)\right]. \end{aligned} \tag{2.47}$$

To obtain Equality (a), we have used Dominated Convergence Theorem (DCT); We employed Lemma 2.5.5 part (vi) to confirm the conditions of DCT. Our goal is to show that the two terms in (2.47) are both negative. Regarding the first term, we first evaluate it by conditioning on $B = b$ for a given constant $b > 0$ (note that B and Z are independent):

$$
\begin{aligned}
&\mathbb{E}_Z\left[(\eta_q(b/\sigma + Z;\chi) - b/\sigma - Z)(\partial_1\eta_q(b/\sigma + Z;\chi) - 1)\right] \\
&\overset{(b)}{=} \int_{-\infty}^{+\infty} (\eta_q(z;\chi) - z)(\partial_1\eta_q(z;\chi) - 1)\phi(z - b/\sigma)dz \\
&= \int_{0}^{\infty} (\eta_q(z;\chi) - z)(\partial_1\eta_q(z;\chi) - 1)\phi(z - b/\sigma)dz + \\
&\quad \int_{-\infty}^{0} (\eta_q(z;\chi) - z)(\partial_1\eta_q(z;\chi) - 1)\phi(z - b/\sigma)dz \\
&\overset{(c)}{=} \int_{0}^{\infty} (\eta_q(z;\chi) - z)(\partial_1\eta_q(z;\chi) - 1)(\phi(z - b/\sigma) - \phi(z + b/\sigma))dz \overset{(d)}{>} 0,
\end{aligned}
$$

where $\phi(\cdot)$ is the density function of standard normal; (b) is obtained by a change of variables; (c) is due to the fact $\partial_1\eta_q(-z;\chi) = \partial_1\eta_q(z;\chi)$ implied by Lemma 2.5.5 part (iv); (d) is based on the following arguments: According to Lemmas 2.5.5 part (ii) and Lemma 2.5.8 part (iii), $\eta_q(z;\chi) < z$ and $\partial_1\eta_q(z;\chi) < 1$ for $z > 0$. Moreover, $\phi(z - b/\sigma) - \phi(z + b/\sigma) > 0$ for $z, b/\sigma > 0$. Hence we have

$$
\mathbb{E}_Z\left[(\eta_q(b/\sigma + Z;\chi) - b/\sigma - Z)(\partial_1\eta_q(b/\sigma + Z;\chi) - 1)(-b/\sigma^2)\right] < 0.
$$

Similarly we can show the above inequality holds for $b < 0$. It is clear that the term on the left hand side equals zero when $b = 0$. Thus we have proved the first term in (2.47) is negative. Now we should discuss the second term. Again we condition on $B = b$ for a given $b > 0$:

$$
\begin{aligned}
\mathbb{E}_Z\left[Z(\partial_1\eta_q(b/\sigma + Z;\chi) - 1)\right] &\overset{(e)}{=} \mathbb{E}(\partial_1^2\eta_q(b/\sigma + Z;\chi)) \\
&= \int_0^{\infty} [\partial_1^2\eta_q(z;\chi)(\phi(z - b/\sigma) - \phi(z + b/\sigma))]dz > 0. \qquad (2.48)
\end{aligned}
$$

Equality (e) is the result of Stein's lemma, i.e. Lemma 2.5.12. Note that the weak differentiability condition required in Stein's lemma is guaranteed by Lemma 2.5.7. To obtain the last inequality, we have used Lemma 2.5.8 part

(iv) and the fact that $\phi(z - b/\sigma) - \phi(z + b/\sigma) > 0$ for $z, b/\sigma > 0$. Hence we obtain that

$$\mathbb{E}_Z\left[Z(\partial_1\eta_q(b/\sigma + Z;\tau) - 1)(-b/\sigma^2)\right] < 0.$$

The same approach would work for $b < 0$, and clearly the left hand side term of the above inequality equals zero for $b = 0$. We can therefore conclude the second term in (2.47) is negative as well.

(iv) $3/2 < q < 2$: The proof of this case is similar to the last one. The only difference is that the proof steps we presented in (2.48) may not work, due to the non-differentiability of $\partial_1\eta_q(u;\chi)$ for $q > 3/2$ as shown in Lemma 2.5.7. Our goal here is to use an alternative approach to prove: $\mathbb{E}_Z\left[Z(\partial_1\eta_q(b/\sigma + Z;\chi) - 1)\right] > 0$ for $b > 0$. We have

$$\begin{aligned}\mathbb{E}_Z\left[Z(\partial_1\eta_q(b/\sigma + Z;\chi) - 1)\right] &= \int_{-\infty}^{\infty} z(\partial_1\eta_q(b/\sigma + z;\chi) - 1)\phi(z)dz \\ &= \int_0^{\infty} z(\partial_1\eta_q(b/\sigma + z;\chi) - \partial_1\eta_q(b/\sigma - z;\chi))\phi(z)dz \\ &= \int_0^{\infty} z(\partial_1\eta_q(|b/\sigma + z|;\chi) - \partial_1\eta_q(|b/\sigma - z|;\chi))\phi(z)dz, \end{aligned} \tag{2.49}$$

where the last equality is due to the fact $\partial_1\eta_q(u;\chi) = \partial_1\eta_q(|u|;\chi)$ for any $u \in \mathbb{R}$. Since $|b/\sigma - z| < |b/\sigma + z|$ for $z, b/\sigma > 0$ and according to Lemma 2.5.8 part (iv), we obtain

$$\partial_1\eta_q(|b/\sigma + z|;\chi) - \partial_1\eta_q(|b/\sigma - z|;\chi) > 0. \tag{2.50}$$

Combining (2.49) and (2.50) completes the proof.

□

Lemma 2.5.13 paves our way in the study of the solution of (2.5). Define

$$\chi_{\min} = \inf\left\{\chi \geq 0 : \frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi)) \leq 1\right\}, \tag{2.51}$$

where $Z \sim N(0,1)$. The following corollary is a conclusion from Lemma 2.5.13.

Corollary 2.5.14. *For a given* $1 \leq q \leq 2$*, Equation* (2.5)*:*

$$\sigma^2 = \sigma_\omega^2 + \frac{1}{\delta}\mathbb{E}_{B,Z}[(\eta_q(B+\sigma Z;\chi\sigma^{2-q}) - B)^2], \quad \sigma_w > 0$$

has a unique solution $\sigma = \sigma_\chi$ *for any* $\chi \in (\chi_{\min}, \infty)$*, and does not have any solution if* $\chi \in (0, \chi_{\min})$*.*

Proof. First note that since $\sigma_w > 0$, $\sigma = 0$ is not a solution of (2.5). Hence we can equivalently write Equation (2.5) in the following form:

$$1 = \frac{\sigma_w^2}{\sigma^2} + \frac{1}{\delta}R_q(\chi, \sigma) \triangleq F(\sigma, \chi). \tag{2.52}$$

According to Lemma 2.5.13, $F(\sigma, \chi)$ is a strictly decreasing function of σ over $(0, \infty)$. We also know that $F(\sigma, \chi)$ is a continuous function of σ from the proof of Lemma 2.5.13. Moreover, it is straightforward to confirm that

$$\lim_{\sigma \to 0} F(\sigma, \chi) = \infty, \quad \lim_{\sigma \to \infty} F(\sigma, \chi) = \frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi)). \tag{2.53}$$

Thus Equation (2.5) has a solution (the uniqueness is automatically guaranteed by the monotonicity of $F(\sigma, \chi)$) if and only if $\frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi)) < 1$. Recall the definition of $\chi_{\min}$ given in (2.51). Since $\mathbb{E}(\eta_q^2(Z;\chi))$ is a strictly decreasing and continuous function of χ, $\frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi)) < 1$ holds if $\chi \in (\chi_{\min}, \infty)$ and fails when $\chi \in (0, \chi_{\min})$. □

Corollary 2.5.14 characterizes the existence and uniqueness of solution for Equation (2.5). Our next goal is to prove that (2.5) and (2.6) share at least one common solution. Our strategy is: among all the pairs (σ_χ, χ) that satisfy (2.5), we show that at least one of them satisfies (2.6). We do this in the next few lemmas.

Lemma 2.5.15. *Let* $\delta < 1$*. For each value of* $\chi \in (\chi_{min}, \infty)$*, define* σ_χ *as the value of* σ *that satisfies* (2.5)*. Then,*

$$\begin{aligned} \lim_{\chi \to \infty} \chi\sigma_\chi^{2-q}\left(1 - \frac{1}{\delta}\mathbb{E}[\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})]\right) &= \infty, \\ \lim_{\chi \to \chi_{\min}^+} \chi\sigma_\chi^{2-q}\left(1 - \frac{1}{\delta}\mathbb{E}[\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})]\right) &= -\infty. \end{aligned} \tag{2.54}$$

Proof. We first show that

$$\lim_{\chi \to \infty} \sigma_\chi^2 = \sigma_w^2 + \frac{\mathbb{E}|B|^2}{\delta}, \qquad \lim_{\chi \to \chi_{\min}^+} \sigma_\chi = \infty. \tag{2.55}$$

For the first part, we only need to show $\sigma_{\chi_k}^2 \to \sigma_w^2 + \frac{\mathbb{E}|B|^2}{\delta}$ for any sequence $\chi_k \to \infty$. For that purpose, we first prove $\sigma_\chi = O(1)$. Otherwise, there exists a sequence $\chi_n \to \infty$ such that $\sigma_{\chi_n} \to \infty$. Because

$$(\eta_q(B/\sigma_{\chi_n} + Z; \chi_n) - B/\sigma_{\chi_n})^2 \leq 2(B/\sigma_{\chi_n} + Z)^2 + 2B^2/\sigma_{\chi_n}^2 \leq 6B^2 + 4Z^2$$

for large enough n, we can apply Dominated Convergence Theorem (DCT) to conclude

$$\lim_{n \to \infty} R_q(\chi_n, \sigma_{\chi_n}) = \mathbb{E} \lim_{n \to \infty} [\eta_q(B/\sigma_{\chi_n} + Z; \chi_n) - B/\sigma_{\chi_n}]^2 = 0.$$

On the other hand, since the pair $(\sigma_{\chi_n}, \chi_n)$ satisfies (2.5) we obtain

$$\lim_{n \to \infty} R_q(\chi_n, \sigma_{\chi_n}) = \lim_{n \to \infty} \delta\big(1 - \frac{\sigma_w^2}{\sigma_{\chi_n}^2}\big) = \delta.$$

This is a contradiction. We next consider any convergent subsequence $\{\sigma_{\chi_{k_n}}\}$ of $\{\sigma_{\chi_k}\}$. The facts $\sigma_{x_k} \geq \sigma_w$ and $\sigma_{\chi_k} = O(1)$ imply $\sigma_{\chi_{k_n}} \to \sigma^* \in (0, \infty)$. Moreover, since

$$(\eta_q(B + \sigma_{\chi_{k_n}} Z; \chi_{k_n} \sigma_{\chi_{k_n}}^{2-q}) - B)^2 \leq 6B^2 + 5(\sigma^*)^2 Z^2,$$

when n is large enough. We can apply DCT to obtain,

$$(\sigma^*)^2 = \lim_{n \to \infty} \sigma_{\chi_{k_n}}^2 = \sigma_w^2 + \frac{1}{\delta} \mathbb{E} \lim_{n \to \infty} (\eta_q(B + \sigma_{\chi_{k_n}} Z; \chi_{k_n} \sigma_{\chi_{k_n}}^{2-q}) - B)^2 = \sigma_w^2 + \frac{\mathbb{E}B^2}{\delta}.$$

Thus, we have showed any convergent subsequence of $\{\sigma_{\chi_k}^2\}$ converges to the same limit $\sigma_w^2 + \frac{\mathbb{E}B^2}{\delta}$. Hence the sequence converges to that limit as well.

Regarding the second part in (2.55), if it is not the case, then there exists a sequence $\chi_n \to \chi_{\min}^+$ such that $\sigma_{\chi_n} = O(1)$. Equation (2.52) shows,

$$1 = \frac{\sigma_w^2}{\sigma_{\chi_n}^2} + \frac{1}{\delta} R_q(\chi_n, \sigma_{\chi_n}) \overset{(a)}{\geq} \frac{\sigma_w^2}{\sigma_{\chi_n}^2} + \frac{1}{\delta} R_q(\chi_n, \infty) = \frac{\sigma_w^2}{\sigma_{\chi_n}^2} + \frac{1}{\delta} \mathbb{E}(\eta_q^2(Z; \chi_n)),$$

where (a) is due to Lemma 2.5.13. From the definition of $\chi_{\min}$ in (2.51), it is clear that $\frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi_{\min})) = 1$ when $\delta < 1$. Hence letting $n \to \infty$ on the both sides of the above inequaitiy leads to $1 \geq \Omega(1) + 1$, which is a contradiction.

We are in position to derive the two limiting results in (2.54). To obtain the first one, note that $\sigma_\chi^2 \to \sigma_w^2 + \frac{\mathbb{E}|B|^2}{\delta}$, as $\chi \to \infty$. Therefore, Lemma 2.5.8 part (vi) combined with DCT gives us

$$\lim_{\chi\to\infty} \mathbb{E}\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q}) = 0.$$

The first result of (2.54) can then be trivially derived. Regarding the second result, we have showed that as $\chi \to \chi_{\min}^+$, $\sigma_\chi \to \infty$. We also have

$$\begin{aligned}\mathbb{E}\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q}) &\overset{(b)}{=} \frac{1}{\sigma_\chi}\mathbb{E}(Z\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})) \\ &\overset{(c)}{=} \mathbb{E}(Z\eta_q(B/\sigma_\chi + Z;\chi)),\end{aligned}$$

where (b) holds by Lemma 2.5.12 and (c) is due to Lemma 2.5.5 part (v). Hence

$$\begin{aligned}&\lim_{\chi\to\chi_{\min}^+} \mathbb{E}\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q}) \\ = &\lim_{\chi\to\chi_{\min}^+} \mathbb{E}(Z\eta_q(B/\sigma_\chi+Z;\chi)) = \mathbb{E}(Z\eta_q(Z;\chi_{\min})) \\ \overset{(d)}{=} &\ \mathbb{E}(\eta_q^2(Z;\chi_{\min})) + \chi_{\min}q\mathbb{E}(|\eta_q(Z;\chi_{\min})|^q) \\ = &\ \delta + \chi_{\min}q\mathbb{E}(|\eta_q(Z;\chi_{\min})|^q),\end{aligned}$$

where (d) is the result of Lemma 2.5.5 part (i). We thus obtain

$$\lim_{\chi\to\chi_{min}^+}\left(1-\frac{1}{\delta}\mathbb{E}\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})\right) = -\frac{1}{\delta}\chi_{\min}q\mathbb{E}|\eta_q(Z;\chi_{\min})|^q. \tag{2.56}$$

Combining (2.56) and the fact that $\chi_{\min} > 0, \sigma_\chi \to \infty$ finishes the proof. □

Lemma 2.5.16. *Let $\delta \geq 1$. For each value of $\chi \in (\chi_{\min},\infty)$, define σ_χ as the value of σ that satisfies* (2.5). *Then,*

$$\begin{aligned}\lim_{\chi\to\infty}\chi\sigma_\chi^{2-q}\left(1-\frac{1}{\delta}\mathbb{E}[\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})]\right) &= \infty, \qquad (2.57)\\ \lim_{\chi\to\chi_{\min}^+}\chi\sigma_\chi^{2-q}\left(1-\frac{1}{\delta}\mathbb{E}[\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})]\right) &= 0.\end{aligned}$$

Proof. The exactly same arguments presented in the proof of Lemma 2.5.15 can be applied to prove the first result in (2.57). We now focus on the proof of the second one. Since $\mathbb{E}|\partial_1\eta_q(B+\sigma_\chi Z;\chi\sigma_\chi^{2-q})|\leq 1$, our goal will be to show $\chi\sigma_\chi^{2-q}=o(1)$, as $\chi\to 0^+$ (note that $\chi_{\min}=0$ when $\delta\geq 1$).

We first consider the case $\delta>1$. To prove $\chi\sigma_\chi^{2-q}=o(1)$, it is sufficient to show $\sigma_\chi=O(1)$. Suppose this is not true, then there exists a sequence $\chi_n\to 0$ such that $\sigma_{\chi_n}\to\infty$. Recall that (χ_n,σ_{χ_n}) satisfies (2.5):

$$\sigma_{\chi_n}^2=\sigma_w^2+\frac{1}{\delta}\mathbb{E}(\eta_q(B+\sigma_{\chi_n}Z;\chi_n\sigma_{\chi_n}^{2-q})-B)^2. \tag{2.58}$$

Dividing both sides of the above equation by $\sigma_{\chi_n}^2$ and letting $n\to\infty$ yields $1=\frac{1}{\delta}<1$, which is a contradiction.

Regarding the case $\delta=1$, we first claim that $\sigma_\chi\to\infty$, as $\chi\to 0$. Otherwise, there exists a sequence $\chi_n\to 0$ such that $\sigma_{\chi_n}\to\sigma^*\in(0,\infty)$. However, taking the limit $n\to\infty$ on both sides of (2.58) gives us $(\sigma^*)^2=\sigma_w^2+(\sigma^*)^2$ where contradiction arises. Hence, if we can show $\chi\sigma_\chi^2=O(1)$, then $\chi\sigma_\chi^{2-q}=o(1)$ will be proved. Starting from (2.58) (replacing χ_n by χ) with $\delta=1$, we can have for $q\in(1,2]$

$$\begin{aligned}
0 &= \sigma_w^2+\sigma_\chi^2\mathbb{E}(\eta_q(B/\sigma_\chi+Z;\chi)-B/\sigma_\chi-Z)^2+\\
&\quad 2\sigma_\chi^2\mathbb{E}Z(\eta_q(B/\sigma_\chi+Z;\chi)-B/\sigma_\chi-Z)\\
&\overset{(a)}{=} \sigma_w^2+\chi\sigma_\chi^2\cdot\underbrace{\mathbb{E}(\chi q^2|\eta_q(B/\sigma_\chi+Z;\chi)|^{2q-2})}_{A}+\\
&\quad \chi\sigma_\chi^2\cdot\underbrace{\mathbb{E}\frac{-2q(q-1)|\eta_q(B/\sigma_\chi+Z;\chi)|^{q-2}}{1+\chi q(q-1)|\eta_q(B/\sigma_\chi+Z;\chi)|^{q-2}}}_{B},
\end{aligned}$$

where to obtain (a) we have used Lemma 2.5.5 part (i), Lemma 2.5.8 part (i) and Lemma 2.5.12. Therefore we obtain

$$\chi\sigma_\chi^2=-\sigma_w^2\cdot(A+B)^{-1}. \tag{2.59}$$

Because $\sigma_\chi\to\infty$ as $\chi\to 0$, it is easily seen that

$$\lim_{\chi\to 0^+}A=0,\quad \liminf_{\chi\to 0^+}|B|\geq 2q(q-1)\mathbb{E}|Z|^{q-2}. \tag{2.60}$$

Combining results (2.59) and (2.60) we can conclude that $\chi\sigma_\chi^2 = O(1)$. Finally for the case $q = 1$, we do similar calculations and have

$$|A| = \frac{1}{\chi}\mathbb{E}(\eta_q(B/\sigma_\chi + Z; \chi) - B/\sigma_\chi - Z)^2 \leq \chi \to 0,$$
$$|B| = \frac{2}{\chi}P(|B/\sigma_\chi + Z| \leq \chi) \to \frac{4}{\sqrt{2\pi}}.$$

This completes the proof. □

According to the results presented in Lemmas 2.5.15 and 2.5.16, if the function $\chi\sigma_\chi^{2-q}\big(1 - \frac{1}{\delta}\mathbb{E}[\partial_1\eta_q(B + \sigma_\chi Z; \chi\sigma_\chi^{2-q})]\big)$, is continuous with respect to $\chi \in (\chi_{\min}, \infty)$, then we can conclude that Equations (2.5) and (2.6) share at least one common solution pair. To confirm the continuity, it is straightforward to employ implicit function theorem to show σ_χ is continuous about χ. Moreover, According to Lemma 2.5.6, $\partial_1\eta_q(u; \chi)$ is also a continuous function of its arguments.

The proof of uniqueness is motivated by the idea presented in Bayati and Montanari (2012). Suppose there are two different solutions denoted by $(\sigma_{\chi_1}, \chi_1)$ and $(\sigma_{\chi_2}, \chi_2)$, respectively. By applying Theorem 2.2.2 (in the next section we will prove the result of Theorem 2.2.2 holds for any solution pair) with $\psi(a, b) = (a - b)^2$, we have

$$\begin{aligned} \text{AMSE}(\lambda, q, \sigma_w) &= \mathbb{E}[\eta_q(B + \sigma_{\chi_1} Z; \chi_1\sigma_{\chi_1}^{2-q}) - B]^2 \\ &\overset{(a)}{=} \delta(\sigma_{\chi_1}^2 - \sigma_w^2), \end{aligned}$$

where (a) is due to (2.5). The same equations hold for the other solution pair $(\sigma_{\chi_2}, \chi_2)$. Since they have the same AMSE, it follows that $\sigma_{\chi_1} = \sigma_{\chi_2}$. Next we choose a different pseudo-Lipschitz function $\psi(a, b) = |a|$ in Theorem 2.2.2 to obtain

$$\begin{aligned} &\lim_{p\to\infty}\frac{1}{p}\sum_{i=1}^{p}|\hat{\beta}_i(\lambda, q, p)| = \mathbb{E}|\eta_q(B + \sigma_{\chi_1} Z; \chi_1\sigma_{\chi_1}^{2-q})| \\ = \;& \mathbb{E}|\eta_q(B + \sigma_{\chi_2} Z; \chi_2\sigma_{\chi_2}^{2-q})| = \mathbb{E}|\eta_q(B + \sigma_{\chi_1} Z; \chi_2\sigma_{\chi_1}^{2-q})| \end{aligned}$$

Since $\mathbb{E}|\eta_q(B + \sigma_{\chi_1} Z; \chi)|$, as a function of $\chi \in (0, \infty)$, is strictly decreasing based on Lemma 2.5.8 part (v), we conclude $\chi_1 = \chi_2$.

2.5.4.3 Approximate message passing algorithms

For a function $f : \mathbb{R} \to \mathbb{R}$ and a vector $v \in \mathbb{R}^m$, we use $f(v) \in \mathbb{R}^m$ to denote the vector $(f(v_1), \ldots, f(v_m))$. Recall $\eta_q(u; \chi)$ is the proximal operator for the function $\| \cdot \|_q^q$. We are in the linear regression model setting: $y = X\beta + w$. To estimate β, we adapt the AMP algorithm in Maleki (2010) to generate a sequence of estimates $\beta^t \in \mathbb{R}^p$, based on the following iterations (initialized at $\beta^0 = 0, z^0 = y$):

$$\begin{aligned} \beta^{t+1} &= \eta_q(X^T z^t + \beta^t; \theta_t), \\ z^t &= y - X\beta^t + \frac{1}{\delta} z^{t-1} \langle \partial_1 \eta_q(X^T z^{t-1} + \beta^{t-1}; \theta_{t-1}) \rangle, \end{aligned} \tag{2.61}$$

where $\langle v \rangle = \frac{1}{p} \sum_{i=1}^p v_i$ denotes the average of a vector's components and $\{\theta_t\}$ is a sequence of tuning parameters specified during the iterations. A remarkable phenomenon about AMP is that the asymptotics of the sequence $\{\beta^t\}$ can be characterized by one dimensional parameter, known as the state of the system. The following theorem clarifies this claim.

Theorem 2.5.17. *Let $\{\beta(p), X(p), w(p)\}$ be a converging sequence and $\psi : \mathbb{R}^2 \to \mathbb{R}$ be a pseudo-Lipschitz function. For any iteration number $t > 0$,*

$$\lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^p \psi(\beta_i^{t+1}, \beta_i) = \mathbb{E}[\psi(\eta_q(B + \tau_t Z; \theta_t), B)], \quad a.s.,$$

where $B \sim f_\beta$ and $Z \sim N(0,1)$ are independent and $\{\tau_t\}_{t=0}^\infty$ can be tracked through the following recursion ($\tau_0^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}|B|^2$):

$$\tau_{t+1}^2 = \sigma_w^2 + \frac{1}{\delta}\mathbb{E}[\eta_q(B + \tau_t Z; \theta_t) - B]^2, \quad t \geq 0. \tag{2.62}$$

Proof. According to Lemma 2.5.5 part (vi), $\eta_q(u; \chi)$ is a Lipschitz continuous function of u. We can then directly apply Theorem 1 in Bayati and Montanari (2011) to complete the proof. □

Equation (2.62) is called state evolution. Theorem 2.5.17 demonstrates that the general asymptotic performance of $\{\beta^t\}$ is sharply predicted by the state evolution.

From now on, we will consider the AMP estimates $\{\beta^t\}$ with $\theta_t = \chi\tau_t^{2-q}$ in (2.61). The positive constant χ is the solution of (2.5) and (2.6). Note we have proved in Chapter 2.5.4.2 that the solution exists. We next present a useful lemma that characterizes the convergence of $\{\tau_t\}$. Recall the definition in (2.51):

$$\chi_{\min} = \inf\left\{\chi \geq 0 : \frac{1}{\delta}\mathbb{E}(\eta_q^2(Z;\chi)) \leq 1\right\}.$$

Lemma 2.5.18. *For any given $\chi \in (\chi_{\min}, \infty)$, the sequence $\{\tau_t\}_{t=0}^{\infty}$ generated from (2.62) with $\theta_t = \chi\tau_t^{2-q}$ converges to a finite number as $t \to \infty$.*

Proof. Denote $\mathcal{H}(\tau) = \sigma_w^2 + \frac{1}{\delta}\mathbb{E}[\eta_q(B + \tau Z; \chi\tau^{2-q}) - B]^2$. According to Corollary 2.5.14, we know $\mathcal{H}(\tau) = \tau^2$ has a unique solution. Furthermore, since $\mathcal{H}(0) > 0$ and $\mathcal{H}(\tau) < \tau^2$ when τ is large enough, it is straightforward to confirm the result stated in the above lemma. □

Denote $\tau_t \to \tau_*$ as $t \to \infty$. Lemma 2.5.18 and (2.62) together yield

$$\tau_*^2 = \sigma_w^2 + \frac{1}{\delta}\mathbb{E}[\eta_q(B + \tau_* Z; \chi\tau_*^{2-q}) - B]^2. \tag{2.63}$$

This is the same as Equation (2.5). We hence see the connection between AMP estimates and bridge regression. The main part of the proof for Theorem 2.2.2 is to rigorously establish such connection. In particular we will show the sequence $\{\beta^t\}$ converges (in certain asymptotic sense) to $\hat{\beta}(\lambda, q)$ as $t \to \infty$ later on. Towards that goal, we present the next theorem that shows asymptotic characterization of other quantities in the AMP algorithm.

Theorem 2.5.19. *Define $w_t \triangleq \frac{1}{\delta}\langle \partial_1\eta_q(X^T z^{t-1} + \beta^{t-1}; \chi\tau_{t-1}^{2-q})\rangle$. Under the conditions of Theorem 2.5.17, we have almost surely*

(i) $\lim\limits_{t\to\infty}\lim\limits_{p\to\infty} \frac{\|\beta^{t+1} - \beta^t\|_2^2}{p} = 0.$

(ii) $\lim\limits_{t\to\infty}\lim\limits_{p\to\infty} \frac{\|z^{t+1} - z^t\|_2^2}{p} = 0.$

(iii) $\lim\limits_{p\to\infty} \frac{\|z^t\|_2^2}{n} = \tau_t^2.$

(iv) $\lim_{p\to\infty} w_t = \frac{1}{\delta}\mathbb{E}[\partial_1 \eta_q(B+\tau_{t-1}Z; \chi\tau_{t-1}^{2-q})]$, *where* B, Z *are the same random variables as in Theorem 2.5.17.*

Proof. All the results for $q = 1$ have been derived in Bayati and Montanari (2012). We here generalize them to the case $1 < q \leq 2$. Since the proof is mostly a direct modification of that in Bayati and Montanari (2012), we only highlight the differences and refer the reader to Bayati and Montanari (2012) for detailed arguments. According to the proof of Lemma 4.3 in Bayati and Montanari (2012), we have almost surely

$$\begin{aligned}
& \lim_{p\to\infty} \frac{\|z^{t+1}-z^t\|_2^2}{p} = \lim_{p\to\infty} \frac{\|\beta^{t+1}-\beta^t\|_2^2}{p} \\
= \ & \mathbb{E}[\eta_q(B+Z_t;\tau_t^{2-q}\chi) - \eta_q(B+Z_{t-1};\tau_{t-1}^{2-q}\chi)]^2,
\end{aligned}$$

where (Z_t, Z_{t-1}) is jointly zero-mean gaussian, independent from $B \sim f_\beta$, with covariance matrix defined by the recursion (4.13) in Bayati and Montanari (2012). From Lemma 2.5.6, we know $\eta_q(u;\chi)$ is a differentiable function over $(-\infty, +\infty) \times (0, \infty)$. Hence we can apply mean value theorem to obtain

$$\begin{aligned}
& \mathbb{E}[\eta_q(B+Z_t;\tau_t^{2-q}\chi) - \eta_q(B+Z_{t-1};\tau_{t-1}^{2-q}\chi)]^2 \\
\leq \ & \mathbb{E}[\partial_1\eta_q(a;b)\cdot(Z_t - Z_{t-1}) + \partial_2\eta_q(a;b)\cdot(\tau_t^{2-q} - \tau_{t-1}^{2-q})\chi]^2 \\
\leq \ & 2\mathbb{E}[(\partial_1\eta_q(a;b))^2\cdot(Z_t-Z_{t-1})^2] + 2\mathbb{E}[(\partial_2\eta_q(a;b))^2\cdot(\tau_t^{2-q}-\tau_{t-1}^{2-q})^2\chi^2] \\
\overset{(a)}{\leq} \ & 2\mathbb{E}[(Z_t-Z_{t-1})^2] + 2(\tau_t^{2-q}-\tau_{t-1}^{2-q})^2\chi^2 q^2\mathbb{E}|a|^{2q-2},
\end{aligned}$$

where (a, b) is a point on a line that connects the two points $(B + Z_t, \tau_t^{2-q}\chi)$ and $(B + Z_{t-1}, \tau_{t-1}^{2-q}\chi)$; we have used Lemma 2.5.5 part (ii) and Lemma 2.5.8 part (i)(ii) to obtain (a). Note that Lemma 2.5.18 implies the second term on the right hand side of the last inequality goes to zero, as $t \to \infty$. Regarding the first term, we can follow similar proof steps as for Lemma 5.7 in Bayati and Montanari (2012) to show $\mathbb{E}(Z_t - Z_{t-1})^2 \to 0$, as $t \to \infty$.

The proof of part (iii) is the same as that of Lemma 4.1 in Bayati and Montanari (2012). We do not repeat the proof here. For (iv), Lemma F.3(b) in Bayati and

Montanari (2012) implies the empirical distribution of $\{((X^T z^{t-1} + \beta^{t-1})_i, \beta_i)\}_{i=1}^p$ converges weakly to the distribution of $(B + \tau_{t-1} Z, B)$. Since the function $J(y, z) \triangleq \partial_1 \eta_q(y; \chi \tau_{t-1}^{2-q})$ is bounded and continuous with respect to (y, z) according to Lemma 2.5.5 part (i) and Lemma 2.5.6, (iv) follows directly from the Portmanteau theorem.

□

2.5.4.4 Two useful theorems

In this section, we refer to two useful theorems that have also been applied and cited in Bayati and Montanari (2012). The first one is regarding the limit of the singular values of random matrices taken from Bai and Yin (1993).

Theorem 2.5.20. *(Bai and Yin, 1993). Let $X \in \mathbb{R}^{n \times p}$ be a matrix having i.i.d. entries with $\mathbb{E}X_{ij} = 0, \mathbb{E}X_{ij}^2 = 1/n$. Denote by $\sigma_{\max}(X), \sigma_{\min}(X)$ the largest and smallest non-zero singular values of X, respectively. If $n/p \to \delta > 0$, as $p \to \infty$, then*

$$\lim_{p \to 0} \sigma_{\max}(X) = \frac{1}{\sqrt{\delta}} + 1, \quad a.s.,$$
$$\lim_{p \to 0} \sigma_{\min}(X) = \left| \frac{1}{\sqrt{\delta}} - 1 \right|, \quad a.s.$$

The second theorem establishes the relation between ℓ_1 and ℓ_2 norm for vectors from random subspace, showed in Kashin (1977).

Theorem 2.5.21. *(Kashin, 1977). For a given constant $0 < v \leq 1$, there exists a universal constant c_v such that for any $p \geq 1$ and a uniformly random subspace V of dimension $p(1-v)$,*

$$\mathbb{P}\Big(\forall \beta \in V : c_v \|\beta\|_2 \leq \frac{1}{\sqrt{p}} \|\beta\|_1\Big) \geq 1 - 2^{-p}.$$

2.5.4.5 The main proof steps

As mentioned before we will use similar arguments as the ones shown in Bayati and Montanari (2012). To avoid redundancy, we will not present all the details and rather emphasize on the differences. We suggest interested readers going over the proof in Bayati and Montanari (2012) before studying this section. Similar to Bayati and Montanari (2012), we start with a lemma that summarizes several structural properties of LQLS formulation. Define $\mathcal{F}(\beta) \triangleq \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_q^q$.

Lemma 2.5.22. *Suppose $\beta, r \in \mathbb{R}^p$ satisfy the following conditions:*

(i) $\|r\|_2 \leq c_1\sqrt{p}$

(ii) $\mathcal{F}(\beta + r) \leq \mathcal{F}(\beta)$

(iii) $\|\nabla F(\beta)\|_2 \leq \sqrt{p}\epsilon$

(iv) $\sup_{0\leq\mu_i\leq 1} \sum_{i=1}^p |\beta_i + \mu_i r_i|^{2-q} \leq pc_2$

(v) $0 < c_3 \leq \sigma_{\min}(X)$*, where* $\sigma_{\min}(X)$ *is defined in Theorem 2.5.20*

(vi) $\|r^{\|}\|_2^2 \leq c_4\frac{\|r^{\|}\|_1^2}{p}$*. The vector* $r^{\|} \in \mathbb{R}^p$ *is the projection of* r *onto ker(X)*[5]

Then there exists a function $f(\epsilon, c_1, c_2, c_3, c_3, c_4, \lambda, q)$ *such that*

$$\|r\|_2 \leq \sqrt{p}f(\epsilon, c_1, c_2, c_3, c_4, \lambda, q).$$

Moreover, $f(\epsilon, c_1, c_2, c_3, c_4, \lambda, q) \to 0$ *as* $\epsilon \to 0$.

Proof. First note that

$$\nabla\mathcal{F}(\beta) = -X^T(y - X\beta) + \lambda q(|\beta_1|^{q-1}\text{sign}(\beta_1), \ldots, |\beta_p|^{q-1}\text{sign}(\beta_p))^T.$$

[5]It is the nullspace of X defined as $\ker(X) = \{\beta \in \mathbb{R}^p \mid X\beta = 0\}$.

Combining it with Condition (ii) we have

$$
\begin{aligned}
0 &\geq \mathcal{F}(\beta + r) - \mathcal{F}(\beta) \\
&= \frac{1}{2}\|y - X\beta - Xr\|_2^2 + \lambda\|\beta + r\|_q^q - \frac{1}{2}\|y - X\beta\|_2^2 - \lambda\|\beta\|_q^q \\
&= \frac{1}{2}\|Xr\|_2^2 - r^T X^T(y - X\beta) + \lambda(\|\beta + r\|_q^q - \|\beta\|_q^q) \\
&= \frac{1}{2}\|Xr\|_2^2 + r^T \nabla\mathcal{F}(\beta) + \lambda\sum_{i=1}^{p}(|\beta_i + r_i|^q - |\beta_i|^q - qr_i|\beta_i|^{q-1}\text{sign}(\beta_i)) \\
&\overset{(a)}{\geq} \frac{1}{2}\|Xr\|_2^2 + r^T\nabla\mathcal{F}(\beta) + \frac{\lambda q(q-1)}{2}\sum_{i=1}^{p}|\beta_i + \mu_i r_i|^{q-2} r_i^2, \qquad (2.64)
\end{aligned}
$$

where (a) is obtained by Lemma 2.5.23 that we will prove shortly and $\{\mu_i\}$ are numbers between 0 and 1. Note that we can decompose r as $r = r^{\|} + r^{\perp}$ such that $r^{\|} \in \ker(X), r^{\perp} \in \ker(X)^{\perp}$. Accordingly Condition (v) yields $c_3^2\|r^{\perp}\|_2^2 \leq \|Xr^{\perp}\|_2^2$. This fact combined with Inequality (2.64) implies

$$
\frac{c_3^2}{2}\|r^{\perp}\|_2^2 \leq \frac{1}{2}\|Xr^{\perp}\|_2^2 = \frac{1}{2}\|Xr\|_2^2 \leq -r^T\nabla\mathcal{F}(\beta) \leq \|r\|_2 \cdot \|\nabla\mathcal{F}(\beta)\|_2 \leq c_1 p\epsilon,
$$

where the last inequality is derived from Condition (i) and (iii). Hence we can obtain

$$
\|r^{\perp}\|_2^2 \leq \frac{2c_1 p\epsilon}{c_3^2}.
$$

Our next step is to bound $\|r^{\|}\|_2^2$. By Cauchy-Schwarz inequality we know

$$
\begin{aligned}
\sum_{i=1}^{p}|r_i| &= \sum_{i=1}^{p}\sqrt{|\beta_i + \mu_i r_i|^{2-q}} \cdot \sqrt{r_i^2|\beta_i + \mu_i r_i|^{q-2}}, \\
&\leq \sqrt{\sum_{i=1}^{p}|\beta_i + \mu_i r_i|^{2-q}} \cdot \sqrt{\sum_{i=1}^{p} r_i^2|\beta_i + \mu_i r_i|^{q-2}}.
\end{aligned}
$$

So

$$
\sum_{i=1}^{p} r_i^2|\beta_i + \mu_i r_i|^{q-2} \geq \frac{\|r\|_1^2}{\sum_{i=1}^{p}|\beta_i + \mu_i r_i|^{2-q}}. \qquad (2.65)
$$

Combining Inequality (2.64) and (2.65) gives us

$$
\|r\|_1^2 \leq \frac{-2r^T\nabla\mathcal{F}(\beta)}{\lambda q(q-1)} \cdot \sum_{i=1}^{p}|\beta_i + \mu_i r_i|^{2-q} \overset{(b)}{\leq} \frac{2c_1 c_2 \epsilon}{\lambda q(q-1)} p^2,
$$

where we have used Conditions (i), (iii), and (iv) to derive (b). Using the upper bounds we obtained for $\|r\|_1^2$ and $\|r^{\perp}\|_2^2$, together with Condition (vi), it is straightforward to verify the following chains of inequalities

$$\begin{aligned}\|r^{\parallel}\|_2^2 &\leq \frac{c_4}{p}\|r^{\parallel}\|_1^2 \leq \frac{2c_4}{p}(\|r\|_1^2 + \|r^{\perp}\|_1^2) \leq \frac{2c_4}{p}(\|r\|_1^2 + p\|r^{\perp}\|_2^2) \\ &\leq \frac{2c_4}{p}\cdot\Big(\frac{2c_1c_2\epsilon}{\lambda q(q-1)}p^2 + \frac{2c_1\epsilon}{c_3^2}p^2\Big) = \Big(\frac{4c_1c_2c_4}{\lambda q(q-1)} + \frac{4c_1c_4}{c_3^2}\Big)\cdot \epsilon p.\end{aligned}$$

We are finally able to derive

$$\|r\|_2^2 = \|r^{\parallel}\|_2^2 + \|r^{\perp}\|_2^2 \leq \Big(\frac{4c_1c_2c_4}{\lambda q(q-1)} + \frac{4c_1c_4}{c_3^2} + \frac{2c_1}{c_3^2}\Big)\cdot \epsilon p.$$

This completes the proof. □

Note that Lemma 2.5.22 is a non-asymptotic and deterministic result. It sheds light on the behavior of the cost function $\mathcal{F}(\beta)$ around its global minimum. Suppose $\beta + r$ is the global minimizer (a reasonable assumption according to Condition (ii)), and if there is another point β having small function value (indicated by its gradient from Condition (iii)), then the distance $\|r\|_2$ between β and the optimal solution $\beta + r$ should also be small. This interpretation should not sound surprising, since we already know $\mathcal{F}(\beta)$ is a strictly convex function. However, Lemma 2.5.22 enables us to characterize this property in a precise way, which is crucial in the high dimensional asymptotic analysis. Based on Lemma 2.5.22, we will set $\beta + r = \hat{\beta}(\lambda, q), \beta = \beta^t$ and then verify all the conditions in Lemma 2.5.22 to conclude $\|r\|_2 = \|\hat{\beta}(\lambda, q) - \beta^t\|_2$ is small. In particular that small distance will vanish as $t \to \infty$, thus establishing the asymptotic equivalence between $\hat{\beta}(\lambda, q)$ and β^t. We perform the analysis in a sequel of lemmas and Proposition 2.5.27.

Lemma 2.5.23. *Given a constant q satisfying $1 < q \leq 2$, for any $x, r \in \mathbb{R}$, there exists a number $0 \leq \mu \leq 1$ such that*

$$|x + r|^q - |x|^q - rq|x|^{q-1}\text{sign}(x) \geq \frac{q(q-1)}{2}|x + \mu r|^{q-2}r^2. \tag{2.66}$$

Proof. Denote $f_q(x) = |x|^q$. When $q = 2$, since $f_2(x)$ is a smooth function over $(-\infty, +\infty)$, we can apply Taylor's theorem to obtain (2.66). For any $1 < q < 2$, note that $f_q''(0) = \infty$, hence Taylor's theorem is not applicable to all the values of $x \in \mathbb{R}$. We prove the inequality above in separate cases. First observe that if (2.66) holds for any $x > 0, r \in \mathbb{R}$, then it is true for any $x < 0, r \in \mathbb{R}$ as well. It is also straightforward to confirm that when $x = 0$, we can always choose $\mu = 1$ to satisfy Inequality (2.66) for any $r \in \mathbb{R}$. We therefore focus on the case $x > 0, r \in \mathbb{R}$.

a. When $x + r > 0$, since $f_q(x)$ is a smooth function over $(0, \infty)$, we can apply Taylor's theorem to obtain (2.66).

b. If $x+r = 0$, choosing $\mu = 0$, Inequality (2.66) is simplified to $(q-1)x^q \geq \frac{q(q-1)}{2}x^q$, which is clearly valid.

c. When $x + r < 0$, we consider two different scenarios.

   i. First suppose $-x-r \geq x$. We apply (2.66) to the pair $-r-x$ and x. Then we know there exists $0 \leq \tilde{\mu} \leq 1$ such that

$$|x + r|^q - |x|^q \geq \frac{q(q-1)}{2}|\tilde{\mu}(-x-r) + (1-\tilde{\mu})x|^{q-2}(2x+r)^2 \\ -(2x+r)q|x|^{q-1}$$

   It is also straightforward to verify that there is $0 \leq \mu \leq 1$ so that $\mu(x + r) + (1-\mu)x = -\tilde{\mu}(-x-r) - (1-\tilde{\mu})x$. Denote

$$g(y) = \frac{q(q-1)}{2}|\tilde{\mu}(-x-r) + (1-\tilde{\mu})x|^{q-2}y^2 + q|x|^{q-1}y.$$

   If we can show $g(-2x-r) \geq g(r)$, we can obtain the Inequality (2.66). It is easily seen that the quadratic function $g(y)$ achieves global minimum at

$$y_0 = \frac{-1}{q-1}|x|^{q-1}|\tilde{\mu}(-x-r) + (1-\tilde{\mu})x|^{2-q} \leq \frac{-1}{q-1}|x| < -x.$$

   Moreover, note that $-2x-r \geq 0, r < 0$ and they are symmetric around $y = -x$, hence $g(-2x-r) \geq g(r)$.

ii. Consider $0 < -x - r < x$. We again use (2.66) for the pair $-x - -r$ and x to obtain

$$|x+r|^q - |x|^q \geq \frac{q(q-1)}{2}|\tilde{\mu}(x+r) - (1-\tilde{\mu})x|^{q-2}(2x+r)^2$$
$$-(2x+r)q|x|^{q-1} \geq (-2x-r)q|x|^{q-1} + \frac{q(q-1)}{2}|x|^{q-2}(2x+r)^2.$$

Denote $h(y) = \frac{q(q-1)}{2}|x|^{q-2}y^2 + q|x|^{q-1}y$. If we can show $h(-2x-r) \geq h(r)$, Inequality (2.66) will be established with $\mu = 0$. Since $h(x)$ achieves global minimum at $y_0 = \frac{-1}{q-1}|x| < -x$ and $-2x - r > r$, we can get $h(-2x-r) \geq h(r)$.

□

The next lemma is similar to Lemma 3.2 in Bayati and Montanari (2012). The proof is adapted from there.

Lemma 2.5.24. *Let $\{\beta(p), X(p), w(p)\}$ be a converging sequence. Denote the solution of LQLS by $\hat{\beta}(\lambda, q)$, and let $\{\beta^t\}_{t\geq 0}$ be the sequence of estimates generated from the AMP algorithm. There exists a positive constant C s.t.*

$$\lim_{t\to\infty}\lim_{p\to\infty}\frac{\|\beta^t\|_2^2}{p} \leq C, \quad a.s.,$$
$$\limsup_{p\to\infty}\frac{\|\hat{\beta}(\lambda,q)\|^2}{p} \leq C, \quad a.s.$$

Proof. To show the first inequality, according to Theorem 2.5.17 and Lemma 2.5.18, choosing a particular pseudo-Lipschitz function $\psi(x,y) = x^2$ we obtain

$$\lim_{t\to\infty}\lim_{p\to\infty}\frac{\|\beta^t\|_2^2}{p} = \mathbb{E}_{B,Z}[\eta_q(B+\tau_* Z; \chi\tau_*^{2-q})]^2 < \infty, \quad a.s.,$$

where $B \sim f_\beta$ and $Z \sim N(0,1)$ are independent. For the second inequality, first note that since $\hat{\beta}(\lambda, q)$ is the optimal solution we have

$$\lambda\|\hat{\beta}(\lambda,q)\|_q^q \leq \mathcal{F}(\hat{\beta}(\lambda,q)) \leq \mathcal{F}(0) = \frac{1}{2}\|y\|_2^2$$
$$= \frac{1}{2}\|X\beta + w\|_2^2 \leq \|X\beta\|_2^2 + \|w\|_2^2 \leq \sigma_{\max}^2(X)\|\beta\|_2^2 + \|w\|_2^2. \tag{2.67}$$

We then consider the decomposition $\hat{\beta}(\lambda, q) = \hat{\beta}(\lambda, q)^{\perp} + \hat{\beta}(\lambda, q)^{\parallel}$, where $\hat{\beta}(\lambda, q)^{\perp} \in \ker(X)^{\perp}$ and $\hat{\beta}(\lambda, q)^{\parallel} \in \ker(X)$. Since $\ker(X)$ is a uniformly random subspace with dimension $p(1-\delta)_+$, we can apply Theorem 2.5.21 to conclude that, there exists a constant $c(\delta) > 0$ depending on δ such that the following holds with high probability,

$$\begin{aligned}\|\hat{\beta}(\lambda, q)\|_2^2 &= \|\hat{\beta}(\lambda, q)^{\parallel}\|_2^2 + \|\hat{\beta}(\lambda, q)^{\perp}\|_2^2 \leq \frac{\|\hat{\beta}(\lambda, q)^{\parallel}\|_1^2}{c(\delta)p} + \|\hat{\beta}(\lambda, q)^{\perp}\|_2^2 \\ &\leq \frac{2\|\hat{\beta}(\lambda, q)^{\perp}\|_1^2 + 2\|\hat{\beta}(\lambda, q)\|_1^2}{c(\delta)p} + \|\hat{\beta}(\lambda, q)^{\perp}\|_2^2 \\ &\leq \frac{2\|\hat{\beta}(\lambda, q)\|_1^2}{c(\delta)p} + \frac{2+c(\delta)}{c(\delta)}\|\hat{\beta}(\lambda, q)^{\perp}\|_2^2. \end{aligned} \tag{2.68}$$

Moreover, Hölder's inequality combined with Inequality (2.67) yields

$$\frac{\|\hat{\beta}(\lambda, q)\|_1}{p} \leq \left(\frac{\|\hat{\beta}(\lambda, q)\|_q^q}{p}\right)^{1/q} \leq \left(\frac{\sigma_{\max}^2(X)\|\beta\|_2^2 + \|w\|_2^2}{\lambda p}\right)^{1/q}. \tag{2.69}$$

Using the results from (2.68) and (2.69), we can then upper bound $\|\hat{\beta}(\lambda, q)\|_2^2$:

$$\frac{\|\hat{\beta}(\lambda, q)\|_2^2}{p} \overset{(a)}{\leq} \frac{2}{c(\delta)}\left(\frac{\sigma_{\max}^2(X)\|\beta\|_2^2 + \|w\|_2^2}{\lambda p}\right)^{2/q} + \frac{2+c(\delta)}{pc(\delta)\sigma_{\min}^2(X)}\|X\hat{\beta}(\lambda, q)^{\perp}\|_2^2.$$

To obtain (a) we have used the fact $\|X\hat{\beta}(\lambda, q)^{\perp}\|_2^2 \geq \sigma_{\min}^2(X)\|\hat{\beta}(\lambda, q)^{\perp}\|_2^2$. We can further bound

$$\|X\hat{\beta}(\lambda, q)^{\perp}\|_2^2 \overset{(b)}{\leq} 2\|y - X\hat{\beta}(\lambda, q)\|_2^2 + 2\|y\|_2^2 \overset{(c)}{\leq} 4\|y\|_2^2 \overset{(d)}{\leq} 8\sigma_{\max}^2(X)\|\beta\|_2^2 + 8\|w\|_2^2.$$

(b) is due to the simple fact $X\hat{\beta}(\lambda, q)^{\perp} = X\hat{\beta}(\lambda, q)$; (c) and (d) hold since $\|y - X\hat{\beta}(\lambda, q)\|_2^2 \leq 2\mathcal{F}(\hat{\beta}(\lambda, q))$ and inequalities in (2.67). Combining the last two chains of inequalities we obtain with probability larger than $1 - 2^{-p}$

$$\begin{aligned}\frac{\|\hat{\beta}(\lambda, q)\|_2^2}{p} \leq{}& \frac{2}{c(\delta)}\left(\frac{\sigma_{\max}^2(X)\|\beta\|_2^2 + \|w\|_2^2}{\lambda p}\right)^{2/q} \\ &+ \frac{16+8c(\delta)}{c(\delta)\sigma_{\min}^2(X)} \cdot \frac{\sigma_{\max}^2(X)\|\beta\|_2^2 + \|w\|_2^2}{p}.\end{aligned}$$

Finally, because both $\sigma_{\min}(X)$ and $\sigma_{\max}(X)$ converge to non-zero constants by Theorem 2.5.20 and (β, X, w) is a converging sequence, the right hand side of the above inequality converges to a finite number. □

Lemma 2.5.25. *Let $\{\beta(p), X(p), w(p)\}$ be a converging sequence. Denote the solution of LQLS by $\hat{\beta}(\lambda, q)$, and let $\{\beta^t\}_{t\geq 0}$ be the sequence of estimates generated from the AMP algorithm.There exists a positive constant $\tilde{C}$ s.t.*

$$\limsup_{t\to\infty}\limsup_{p\to\infty}\sup_{0\leq\mu_i\leq 1}\frac{\sum_{i=1}^p|\mu_i\hat{\beta}_i(\lambda,q)+(1-\mu_i)\beta_i^t|^{2-q}}{p}<\tilde{C}, \quad a.s.$$

Proof. For any given $0\leq\mu_i\leq 1$, it is straightforward to see

$$|\mu_i\hat{\beta}_i(\lambda,q)+(1-\mu_i)\beta_i^t|^{2-q}\leq\max\{|\hat{\beta}_i(\lambda,q)|^{2-q},|\beta_i^t|^{2-q}\}\leq|\hat{\beta}_i(\lambda,q)|^{2-q}+|\beta_i^t|^{2-q}.$$

Hence using Hӧlder's inequality gives us

$$\begin{aligned}&\sup_{0\leq\mu_i\leq 1}\frac{1}{p}\sum_{i=1}^p|\mu_i\hat{\beta}_i(\lambda,q)+(1-\mu_i)\beta_i^t|^{2-q}\\ \leq\ &\frac{1}{p}\sum_{i=1}^p|\hat{\beta}_i(\lambda,q)|^{2-q}+\frac{1}{p}\sum_{i=1}^p|\beta_i^t|^{2-q}\leq\Big(\frac{1}{p}\sum_{i=1}^p|\hat{\beta}_i(\lambda,q)|^2\Big)^{\frac{2-q}{2}}+\Big(\frac{1}{p}\sum_{i=1}^p|\beta_i^t|^2\Big)^{\frac{2-q}{2}}.\end{aligned}$$

Applying Lemma 2.5.24 to the above inequality finishes the proof. □

The next lemma is similar to Lemma 3.3 in Bayati and Montanari (2012). The proof is adapted from there.

Lemma 2.5.26. *Let $\{\beta(p), X(p), w(p)\}$ be a converging sequence. Let $\{\beta^t\}_{t\geq 0}$ be the sequence of estimates generated from the AMP algorithm. We have*

$$\lim_{t\to\infty}\lim_{p\to\infty}\frac{\|\nabla\mathcal{F}(\beta^t)\|_2^2}{p}=0, \quad a.s.$$

Proof. Recall the AMP updating rule (2.61):

$$\beta^t=\eta_q(X^Tz^{t-1}+\beta^{t-1};\tau_{t-1}^{2-q}\chi).$$

According to Lemma 2.5.5 part (i) we know β^t satisfies

$$X^Tz^{t-1}+\beta^{t-1}=\beta^t+\tau_{t-1}^{2-q}\chi q(|\beta_1^t|^{q-1}\text{sign}(\beta_1^t),\ldots,|\beta_p^t|^{q-1}\text{sign}(\beta_p^t))^T.$$

The rule (2.61) also tells us

$$z^t=y-X\beta^t+w_tz^{t-1},$$

where w_t is defined in Theorem 2.5.19. Note

$$\nabla\mathcal{F}(\beta^t) = -X^T(y - X\beta^t) + \lambda q(|\beta_1^t|^{q-1}\text{sign}(\beta_1^t), \dots, |\beta_p^t|^{q-1}\text{sign}(\beta_p^t))^T.$$

We can then upper bound $\nabla\mathcal{F}(\beta^t)$ in the following way:

$$\begin{aligned}
\frac{1}{\sqrt{p}}\|\nabla\mathcal{F}(\beta^t)\|_2 &= \frac{1}{\sqrt{p}}\| - X^T(y - X\beta^t) + \lambda(\tau_{t-1}^{2-q}\chi)^{-1}X^Tz^{t-1} + \beta^{t-1} - \beta^t\|_2 \\
&= \frac{1}{\sqrt{p}}\| - X^T(z^t - w_t z^{t-1}) + \lambda(\tau_{t-1}^{2-q}\chi)^{-1}(X^Tz^{t-1} + \beta^{t-1} - \beta^t)\|_2 \\
&\le \frac{\lambda\|\beta^{t-1} - \beta^t\|_2}{\tau_{t-1}^{2-q}\chi\sqrt{p}} + \frac{\|X^T(z^{t-1} - z^t)\|_2}{\sqrt{p}} + \frac{|\lambda + \tau_{t-1}^{2-q}\chi(w_t - 1)| \cdot \|X^Tz^{t-1}\|_2}{\tau_{t-1}^{2-q}\chi\sqrt{p}} \\
&\le \frac{\lambda\|\beta^{t-1} - \beta^t\|_2}{\tau_{t-1}^{2-q}\chi\sqrt{p}} + \frac{\sigma_{\max}(X)\|z^{t-1} - z^t\|_2}{\sqrt{p}} + \frac{\sigma_{\max}(X)|\lambda + \tau_{t-1}^{2-q}\chi(w_t - 1)| \cdot \|z^{t-1}\|_2}{\tau_{t-1}^{2-q}\chi\sqrt{p}}.
\end{aligned}$$

By Lemma 2.5.18, Theorem 2.5.19 part (i)(ii) and Theorem 2.5.20, it is straightforward to confirm that the first two terms on the right hand side of the last inequality vanish almost surely, as $p \to \infty, t \to \infty$. For the third term, Lemma 2.5.18 and Theorem 2.5.19 part (iii)(iv) imply

$$\begin{aligned}
&\lim_{t\to\infty}\lim_{p\to\infty} \frac{|\lambda + \tau_{t-1}^{2-q}\chi(w_t - 1)| \cdot \|z^{t-1}\|_2}{\tau_{t-1}^{2-q}\chi\sqrt{p}} \\
= \quad & \frac{\sqrt{\delta}\tau_*}{\tau_*^{2-q}\chi}\left|\lambda - \tau_*^{2-q}\chi\Big(1 - \frac{1}{\delta}\mathbb{E}\eta_q'(B + \tau_* Z; \tau_*^{2-q}\chi)\Big)\right| = 0, \quad a.s.
\end{aligned}$$

To obtain the last equality, we have used Equation (2.6). □

We are in position to prove the asymptotic equivalence between AMP estimates and bridge regression.

Proposition 2.5.27. *Let $\{\beta(p), X(p), w(p)\}$ be a converging sequence. Denote the solution of LQLS by $\hat{\beta}(\lambda, q)$, and let $\{\beta^t\}_{t\ge0}$ be the sequence of estimates generated from the AMP algorithm. We then have*

$$\lim_{t\to\infty}\lim_{p\to\infty}\frac{1}{p}\|\hat{\beta}(\lambda, q) - \beta^t\|_2^2 = 0, \quad a.s. \tag{2.70}$$

Proof. We utilize Lemma 2.5.22. Let $\beta + r = \hat{\beta}(\lambda, q), \beta = \beta^t$. If this pair of β and r satisfies the conditions in Lemma 2.5.22, we will have $\frac{\|r\|_2^2}{p} = \frac{\|\hat{\beta}(\lambda,q)-\beta^t\|_2}{p}$ being very small. In the rest of the proof, we aim to verify that the conditions in Lemma 2.5.22 hold with high probability and establish the connection between the iteration numbers t and ϵ in Lemma 2.5.22.

a. Condition (i) follows from Lemma 2.5.24:

$$\lim_{t\to\infty}\limsup_{p\to\infty}\frac{\|r\|_2}{\sqrt{p}} \leq \limsup_{p\to\infty}\frac{\|\hat{\beta}(\lambda,q)\|_2}{\sqrt{p}} + \lim_{t\to\infty}\lim_{p\to\infty}\frac{\|\beta^t\|_2}{\sqrt{p}} \leq 2\sqrt{C}, \quad a.s.$$

b. Condition (ii) holds since $\hat{\beta}(\lambda, q)$ is the optimal solution of $\mathcal{F}(\beta)$.

c. Condition (iii) holds by Lemma 2.5.26. Note that $\epsilon \to 0$, as $t \to \infty$.

d. Condition (iv) is due to Lemma 2.5.25.

e. Condition (v) is the result of Theorem 2.5.20.

f. Condition (vi) is a direct application of Theorem 2.5.21.

Note all the claims made above hold almost surely as $p \to \infty$; and $\epsilon \to 0$ as $t \to \infty$. Hence the result (2.70) follows. □

Based on the results from Theorem 2.5.17, Lemma 2.5.18 and Proposition 2.5.27, we can use exactly the same arguments as in the proof of Theorem 1.5 from Bayati and Montanari (2012) to finish the proof of Theorem 2.2.2. Since the arguments are straightforward, we do not repeat it here.

2.5.5 Proof of Theorem 2.2.3

First note that

$$\begin{aligned}\frac{1}{p}\mathbb{E}\|\hat{\beta}(\lambda,q) - \beta\|_2^2 &= \frac{1}{p}\mathbb{E}\sum_{j=1}^{N}(\hat{\beta}_j(\lambda,q) - \beta_j)^2 \\ &= \mathbb{E}(\hat{\beta}_1(\lambda,q) - \beta)^2,\end{aligned}$$

where the last equality is due to the symmetry of the problem. Furthermore, note that according to the Replica claim (Rangan et al., 2012), $(\hat{\beta}_1(\lambda, q), \beta_1)$ converges in distribution to $(\eta_q(B + \bar{\sigma}Z; \bar{\chi}\bar{\sigma}^{2-q}), B)$, where $(\bar{\sigma}, \bar{\chi})$ satisfies (2.5) and (2.6). Hence, by employing the continuous mapping theorem (for convergence in distribution), we conclude that $(\hat{\beta}_1(\lambda, q) - \beta_1)^2$ converges to $(\eta_q(B + \bar{\sigma}Z; \bar{\chi}\bar{\sigma}^{2-q}) - B)^2$ in distribution. Note that convergence in distribution implies convergence of expectations if and only if the sequence of random variables is asymptotically uniformly integrable (See Chapter 2 of Van der Vaart (2000) for more information). According to square integrability assumption, $(\hat{\beta}_j(\lambda, q) - \beta_j)^2$ is asymptotically uniformly integrable and hence $\mathbb{E}(\hat{\beta}_j(\lambda, q) - \beta_j)^2 \to \mathbb{E}(\eta_q(B + \bar{\sigma}Z; \bar{\chi}\bar{\sigma}^{2-q}) - B)^2$.

2.5.6 Proof of Corollary 2.2.4

We prove the result for $q \in [1, 2]$. The proof for other cases follows similarly. According to Theorem 2.2.2, the key of proving Corollary 2.2.4 is to analyze the following equations:

$$\sigma^2 = \sigma_\omega^2 + \frac{1}{\delta}\mathbb{E}[(\eta_q(B + \sigma Z; \chi\sigma^{2-q}) - B)^2], \tag{2.71}$$

$$\lambda_{*,q} = \chi\sigma^{2-q}\big(1 - \frac{1}{\delta}\mathbb{E}[\eta_q'(B + \sigma Z; \chi\sigma^{2-q})]\big), \tag{2.72}$$

where $\lambda_{*,q} = \arg\min_{\lambda \geq 0} \text{AMSE}(\lambda, q, \sigma_w)$. We present the main result in the following lemma.

Lemma 2.5.28. *For every $1 \leq q \leq 2$ and a given optimal tuning $\lambda_{*,q}$, there exists a unique solution pair $(\bar{\sigma}, \bar{\chi})$ that satisfies* (2.71) *and* (2.72)*. Furthermore, $\bar{\sigma}$ is the unique solution of*

$$\sigma^2 = \sigma_w^2 + \frac{1}{\delta}\min_{\chi \geq 0} \mathbb{E}(\eta_q(B + \sigma Z; \chi) - B)^2, \tag{2.73}$$

and

$$\bar{\chi} \in \arg\min_{\chi \geq 0} \mathbb{E}[(\eta_q(B + \bar{\sigma}Z; \chi\bar{\sigma}^{2-q}) - B)^2]. \tag{2.74}$$

Proof. The first part of this lemma directly comes from Lemma 2.5.11. We focus on the proof of the second part. Recall the definition of $R_q(\chi, \sigma)$ in (2.46). Define

$$G_q(\sigma) \triangleq \frac{\sigma_w^2}{\sigma^2} + \frac{1}{\delta} \min_{\chi \geq 0} R_q(\chi, \sigma).$$

Then (2.73) is equivalent to $G_q(\sigma) = 1$. We first show that $G_q(\sigma)$ is a strictly decreasing function of σ over $(0, \infty)$. For any given $\sigma_1 > \sigma_2 > 0$, we can choose χ_1, χ_2 such that

$$\begin{aligned} \chi_1 &= \arg\min_{\chi \geq 0} \mathbb{E}(\eta_q(B/\sigma_1 + Z; \chi) - B/\sigma_1)^2, \\ \chi_2 &= \arg\min_{\chi \geq 0} \mathbb{E}(\eta_q(B/\sigma_2 + Z; \chi) - B/\sigma_2)^2. \end{aligned}$$

Applying Lemma 2.5.13 we have

$$R_q(\chi_1, \sigma_1) = \min_{\chi \geq 0} R_q(\chi, \sigma_1) \leq R_q(\chi_2, \sigma_1) \leq R_q(\chi_2, \sigma_2) = \min_{\chi \geq 0} R_q(\chi, \sigma_2).$$

Hence we obtain that $G_q(\sigma_1) < G_q(\sigma_2)$. We next show

$$\lim_{\sigma \to 0} G_q(\sigma) = \infty, \quad \lim_{\sigma \to \infty} G_q(\sigma) < 1. \tag{2.75}$$

The first result in (2.75) is obvious. To prove the second one, we know

$$\lim_{\sigma \to \infty} G_q(\sigma) \leq \lim_{\sigma \to \infty} \frac{\sigma_w^2}{\sigma^2} + \frac{1}{\delta} R_q(\chi, \sigma) = \frac{1}{\delta} \mathbb{E}\eta_q^2(Z; \chi),$$

for any given $\chi \geq 0$. Choosing a sufficiently large χ completes the proof for the second inequality in (2.75). Based on (2.75) and the fact that $G_q(\sigma)$ is a strictly decreasing and continuous function of σ over $(0, \infty)$, we can conclude (2.73) has a unique solution. Call it σ^*, and denote

$$\chi^* \in \arg\min_{\chi \geq 0} \mathbb{E}[(\eta_q(B + \sigma^* Z; \chi(\sigma^*)^{2-q}) - B)^2].$$

Note that χ^* may not be unique. Further define

$$\lambda^* = \chi^*(\sigma^*)^{2-q}\Big(1 - \frac{1}{\delta} \mathbb{E}[\partial_1 \eta_q(B + \sigma^* Z; \chi^*(\sigma^*)^{2-q})]\Big).$$

It is straightforward to see that the pair (σ^*, χ^*) satisfies (2.5) and (2.6) with $\lambda = \lambda^*$. According to Theorem 2.2.2, we obtain

$$\text{AMSE}(\lambda^*, q, \sigma_w) = \delta((\sigma^*)^2 - \sigma_w^2). \tag{2.76}$$

Also we already know for the optimal tuning $\lambda_{*,q}$

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \delta(\bar{\sigma}^2 - \sigma_w^2) \le \text{AMSE}(\lambda^*, q, \sigma_w).$$

Therefore $\bar{\sigma} \le \sigma^*$. On the other hand,

$$\begin{aligned} \bar{\sigma}^2 &= \sigma_w^2 + \frac{1}{\delta}\mathbb{E}(\eta_q(B + \bar{\sigma}Z; \bar{\chi}\bar{\sigma}^{2-q}) - B)^2 \\ &\ge \sigma_w^2 + \frac{1}{\delta}\min_{\chi \ge 0}\mathbb{E}(\eta_q(B + \bar{\sigma}Z; \chi) - B)^2 = \bar{\sigma}^2 \cdot G_q(\bar{\sigma}). \end{aligned}$$

Thus $G_q(\bar{\sigma}) \le 1 = G_q(\sigma^*)$. Since $G_q(\sigma)$ is a strictly decreasing function, we then obtain $\bar{\sigma} \ge \sigma^*$. Consequently, we conclude $\bar{\sigma} = \sigma^*$. Finally we claim (2.74) has to hold. Otherwise,

$$\begin{aligned} &\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \mathbb{E}(\eta_q(B + \bar{\sigma}Z; \bar{\chi}\bar{\sigma}^{2-q}) - B)^2 \\ > \ &\min_{\chi \ge 0}\mathbb{E}(\eta_q(B + \bar{\sigma}Z; \chi) - B)^2 = \mathbb{E}(\eta_q(B + \sigma^* Z; \chi^*) - B)^2 \\ = \ &\text{AMSE}(\lambda^*, q, \sigma_w), \end{aligned}$$

contradicts the fact that $\lambda_{*,q}$ is the optimal tuning. □

Remark: Lemma 2.5.28 leads directly to the result of Corollary 2.2.4. Furthermore, from the proof of Lemma 2.5.28, we see that if $\mathbb{E}[(\eta_q(B + \sigma Z; \chi\sigma^{2-q}) - B)^2]$, as a function of χ, has a unique minimizer for any given $\sigma > 0$, then $\bar{\chi} = \chi^*$. That means the optimal tuning value $\lambda_{*,q}$ is unique. Mousavi et al. (2017) has proved that $\mathbb{E}[(\eta_q(B + \sigma Z; \chi\sigma^{2-q}) - B)^2]$ is quasi-convex and has a unique minimizer for $q = 1$. We conjecture that it is true for $q \in (1, 2]$ as well and leave it for future research.

2.5.7 Proof of Theorem 2.2.5

2.5.7.1 Roadmap of the proof

Different from the result of Theorem 2.2.7 for LASSO that bounds the second order term in $\mathrm{AMSE}(\lambda_{*,1}, 1, \sigma_w)$, Theorem 2.2.5 characterizes the precise analytical expression of the second dominant term for $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$ with $q \in (1, \infty)$. However, the idea of this proof is similar to the one for Theorem 2.2.7, though the detailed proof steps are more involved here. We suggest interested readers first going over the proof of Theorem 2.2.7 in Chapter 2.5.2 and then this section so that both the proof idea and technical details are smoothly understood. Recall the definition we introduced in (2.46):

$$R_q(\chi, \sigma) = \mathbb{E}(\eta_q(B/\sigma + Z; \chi) - B/\sigma)^2,$$

where $Z \sim N(0, 1)$ and B with the distribution $f_\beta(b) = (1 - \epsilon)\delta_0(b) + \epsilon g(b)$ are independent. Define

$$\chi_q^*(\sigma) = \arg\min_{\chi \geq 0} R_q(\chi, \sigma). \tag{2.77}$$

According to Lemma 2.5.6, it is straightforward to show $R_q(\chi, \sigma)$ is a differentiable function of χ. It is also easily seen that

$$\lim_{\chi \to \infty} R_q(\chi, \sigma) = \frac{\mathbb{E}|B|^2}{\sigma^2}, \quad \lim_{\chi \to 0} R_q(\chi, \sigma) = 1.$$

Therefore, the minimizer $\chi_q^*(\sigma)$ exists at least for sufficiently small σ. If it is not unique, we will consider the one having smallest value itself. As like the proof of Theorem 2.2.7, the key is to characterize the convergence rate for $R_q(\chi_q^*(\sigma), \sigma)$ as $\sigma \to 0$. After having that convergence rate result, we can then obtain the convergence rate for $\bar{\sigma}$ from Equation (2.10) and finally derive the expansion of $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$ based on Corollary 2.2.4. We organize our proof steps as follows:

1. We first characterize the convergence rate of $\chi_q^*(\sigma)$ in Chapter 2.5.7.2.

2. We then obtain the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$ in Chapter 2.5.7.3.

3. We finally derive the second-order expansion for AMSE$(\lambda_{*,q}, q, \sigma_w)$ in Chapter 2.5.7.4.

As a final remark, once we show the proof for $q \in (1, 2)$, since $\eta_2(u; \chi) = \frac{u}{1+2\chi}$ has a nice explicit form, the proof for $q = 2$ can be easily derived. We hence skip it for simplicity. The proof for the case $q > 2$ can be significantly simplified due to the specific range of q. We also do not repeat it here. Refer to Wang et al. (2017) for the complete proof.

2.5.7.2 Characterizing the convergence rate of $\chi_q^*(\sigma)$

The goal of this section is to derive the convergence rate of $\chi_q^*(\sigma)$ as $\sigma \to 0$. We will make use of the fact that $\chi = \chi_q^*(\sigma)$ is the minimizer of $R_q(\chi, \sigma)$, to first show $\chi_q^*(\sigma) \to 0$ and then obtain the rate $\chi_q^*(\sigma) \propto \sigma^{2q-2}$. This is done in the following three lemmas.

Lemma 2.5.29. *Let $\chi_q^*(\sigma)$ denote the minimizer of $R_q(\chi, \sigma)$ as defined in* (2.77). *Then for every $b \neq 0$ and $z \in \mathbb{R}$,*

$$|\eta_q(b/\sigma + z; \chi_q^*(\sigma))| \to \infty, \quad \text{as } \sigma \to 0.$$

Proof. Suppose this is not the case. Then there exist a value of $b \neq 0, z \in \mathbb{R}$ and a sequence $\sigma_k \to 0$, such that $|\eta_q(b/\sigma_k + z; \chi_q^*(\sigma_k))|$ is bounded. Combined with Lemma 2.5.5 part (i) we obtain

$$\chi_q^*(\sigma_k) = \frac{|b/\sigma_k + z| - |\eta_q(b/\sigma_k + z; \chi_q^*(\sigma_k))|}{q|\eta_q(b/\sigma_k + z; \chi_q^*(\sigma_k))|^{q-1}} = \Omega\Big(\frac{1}{\sigma_k}\Big). \tag{2.78}$$

We next show that the result (2.78) implies for any other $\tilde{b} \neq 0$ and $\tilde{z} \in \mathbb{R}$, $|\eta_q(\tilde{b}/\sigma_k + \tilde{z}; \chi_q^*(\sigma_k))|$ is bounded as well. From Lemma 2.5.5 part (i) we know

$$|\tilde{b}/\sigma_k + \tilde{z}| = |\eta_q(\tilde{b}/\sigma_k + \tilde{z}; \chi_q^*(\sigma_k))| + \chi_q^*(\sigma_k) q |\eta_q(\tilde{b}/\sigma_k + \tilde{z}; \chi_q^*(\sigma_k))|^{q-1} \tag{2.79}$$

If $|\eta_q(\tilde{b}/\sigma_k+\tilde{z};\chi_q^*(\sigma_k))|$ is unbounded, then the right hand side of equation (2.79) (take a subsequence if necessary) has the order larger than $1/\sigma_k$. Hence (2.79) can not hold for all the values of k. We thus have reached the conclusion that $|\eta_q(b/\sigma_k+z;\chi_q^*(\sigma_k))|$ is bounded for every $b \neq 0$ and $z \in \mathbb{R}$. Therefore,

$$|\eta_q(G/\sigma_k + Z;\chi_q^*(\sigma_k)) - G/\sigma_k| \to \infty \ a.s., \quad \text{as } k \to \infty,$$

where G has the distribution $g(\cdot)$. We then use Fatou's lemma to obtain

$$R_q(\chi_q^*(\sigma_k),\sigma_k) \geq \epsilon \mathbb{E}|\eta_q(G/\sigma_k + Z;\chi_q^*(\sigma_k)) - G/\sigma_k|^2 \to \infty.$$

On the other hand, since $\chi_q^*(\sigma_k)$ minimizes $R_q(\chi,\sigma_k)$,

$$R_q(\chi_q^*(\sigma_k),\sigma_k) \leq R_q(0,\sigma_k) = 1.$$

Such contradiction completes the proof. □

Lemma 2.5.29 enables us to derive $\chi_q^*(\sigma) \to 0$ as $\sigma \to 0$. We present it in the next lemma.

Lemma 2.5.30. *Let $\chi_q^*(\sigma)$ denote the minimizer of $R_q(\chi,\sigma)$ as defined in* (2.77)*. Then $\chi_q^*(\sigma) \to 0$ as $\sigma \to 0$.*

Proof. First note that

$$\begin{aligned} R_q(\chi,\sigma) &= (1-\epsilon)\mathbb{E}(\eta_q(Z;\chi))^2 + \epsilon\mathbb{E}(\eta_q(G/\sigma+Z;\chi) - G/\sigma - Z)^2 + \\ &\quad 2\epsilon\mathbb{E}Z(\eta_q(G/\sigma+Z;\chi) - G/\sigma - Z) + \epsilon \\ &\overset{(a)}{=} (1-\epsilon)\mathbb{E}(\eta_q(Z;\chi))^2 + \epsilon\chi^2 q^2 \mathbb{E}|\eta_q(G/\sigma+Z;\chi)|^{2q-2} + \\ &\quad 2\epsilon\mathbb{E}(\partial_1\eta_q(G/\sigma+Z;\chi) - 1) + \epsilon \\ &\overset{(b)}{=} (1-\epsilon)\mathbb{E}(\eta_q(Z;\chi))^2 + \epsilon\chi^2 q^2 \mathbb{E}|\eta_q(G/\sigma+Z;\chi)|^{2q-2} + \\ &\quad 2\epsilon\mathbb{E}\left(\frac{1}{1+\chi q(q-1)|\eta_q(G/\sigma+Z;\chi)|^{q-2}} - 1\right) + \epsilon. \end{aligned} \tag{2.80}$$

We have employed Lemma 2.5.5 part (i) and Lemma 2.5.6 to obtain (a); (b) is due to Lemma 2.5.8 part (i). According to Lemma 2.5.29, $|\eta_q(G/\sigma+Z;\chi_q^*(\sigma))| \to \infty \ a.s.$, as $\sigma \to 0$. Hence, if $\chi_q^*(\sigma) \nrightarrow 0$, the second term in (2.80) (with $\chi = \chi_q^*(\sigma)$) goes off to infinity, while the other terms remain finite, and consequently $R_q(\chi_q^*(\sigma), \sigma) \to \infty$. This is a contradiction with the fact $R_q(\chi_q^*(\sigma), \sigma) \leq R_q(0, \sigma) = 1$. □

So far we have shown $\chi_q^*(\sigma) \to 0$ as $\sigma \to 0$. Our next lemma further characterizes the convergence rate of $\chi_q^*(\sigma)$.

Lemma 2.5.31. *Suppose* $\mathbb{P}(|G| \leq t) = O(t)$ *(as* $t \to 0$*) and* $\mathbb{E}|G|^2 < \infty$. *Then for* $q \in (1,2)$ *we have*

$$\lim_{\sigma \to 0} \frac{\chi_q^*(\sigma)}{\sigma^{2q-2}} = \frac{(1-\epsilon)\mathbb{E}|Z|^q}{\epsilon q \mathbb{E}|G|^{2q-2}}.$$

Proof. We first claim that $\chi_q^*(\sigma) = \Omega(\sigma^{2q-2})$. Otherwise there exists a sequence $\sigma_k \to 0$ such that $\chi_q^*(\sigma_k) = o(\sigma_k^{2q-2})$. According to Lemma 2.5.32 (we postpone Lemma 2.5.32 since it deals with $R_q(\chi, \sigma)$),

$$\lim_{k \to \infty} \frac{R_q(\chi_q^*(\sigma_k), \sigma_k) - 1}{\sigma_k^{2q-2}} = 0.$$

On the other hand, by choosing $\chi(\sigma_k) = C\sigma_k^{2q-2}$ with $C = \frac{(1-\epsilon)\mathbb{E}|Z|^q}{\epsilon q \mathbb{E}|G|^{2q-2}}$, Lemma 2.5.32 implies that

$$\lim_{k \to \infty} \frac{R_q(\chi(\sigma_k), \sigma_k) - 1}{\sigma_k^{2q-2}} < 0,$$

which contradicts with the fact that $\chi_q^*(\sigma_k)$ is the minimizer of $R_q(\chi, \sigma_k)$. Moreover, this choice of C shows that for sufficiently small σ there exists $\chi(\sigma)$ such that

$$R_q(\chi_q^*(\sigma), \sigma) \leq R_q(\chi(\sigma), \sigma) < R_q(0, \sigma) = 1.$$

That means $\chi_q^*(\sigma)$ is a non-zero finite value. Hence it satisfies $\frac{\partial R_q(\chi_q^*(\sigma), \sigma)}{\partial \chi} = 0$. From now on we use χ^* to denote $\chi_q^*(\sigma)$ for notational simplicity. That equation can be detailed out as follows:

$$\begin{aligned}0 &= (1-\epsilon)\mathbb{E}\eta_q(Z;\chi^*)\partial_2\eta_q(Z;\chi^*) + \epsilon\mathbb{E}(\eta_q(G/\sigma+Z;\chi^*) - G/\sigma)\partial_2\eta_q(G/\sigma+Z;\chi^*) \\ &\overset{(a)}{=} (1-\epsilon)\underbrace{\mathbb{E}\frac{-q|\eta_q(Z;\chi^*)|^q}{1+\chi^* q(q-1)|\eta_q(Z;\chi^*)|^{q-2}}}_{H_1} + \epsilon\underbrace{\mathbb{E}(Z\partial_2\eta_q(G/\sigma+Z;\chi^*))}_{H_2} \\ &\quad + \epsilon\chi^*\underbrace{\mathbb{E}\frac{q^2|\eta_q(G/\sigma+Z;\chi^*)|^{2q-2}}{1+\chi^* q(q-1)|\eta_q(G/\sigma+Z;\chi^*)|^{q-2}}}_{H_3}, \end{aligned} \tag{2.81}$$

where we have used Lemma 2.5.5 part (i) and Lemma 2.5.8 part (ii) to obtain (a). We now analyze the three terms H_1, H_2 and H_3, respectively. According to Lemma 2.5.30, we have that $\eta_q(Z;\chi^*) \to Z$ as $\sigma \to 0$. Lemma 2.5.5 part (ii) enables us to bound the expression inside the expectation of H_1 by $q|Z|^q$. Hence we can employ Dominated Convergence Theorem (DCT) to obtain

$$\lim_{\sigma\to 0} H_1 = -q\mathbb{E}|Z|^q. \tag{2.82}$$

For the term H_3, we first note that

$$\begin{aligned}&\lim_{\sigma\to 0}\eta_q(G+\sigma Z;\sigma^{2-q}\chi^*) = G, \\ &|\eta_q(G+\sigma Z;\sigma^{2-q}\chi^*)| \le |B| + \sigma|Z|.\end{aligned}$$

We also know that $|\eta_q(G/\sigma+Z;\chi^*)| \to \infty, a.s.$ by Lemma 2.5.29. We therefore can apply DCT to conclude

$$\lim_{\sigma\to 0}\frac{H_3}{\sigma^{2-2q}} = \mathbb{E}\lim_{\sigma\to 0}\frac{q^2|\eta_q(G+\sigma Z;\sigma^{2-q}\chi^*)|^{2q-2}}{1+\chi^* q(q-1)|\eta_q(G/\sigma+Z;\chi^*)|^{q-2}} = q^2\mathbb{E}|G|^{2q-2}. \tag{2.83}$$

We now study the remaining term H_2. According to Lemma 2.5.6, $\partial_2\eta_q(G/\sigma+Z;\chi^*)$ is differentiable with respect to its first argument. So we can apply Lemma 2.5.12 to get

$$\begin{aligned} H_2 &= q(1-q)\underbrace{\mathbb{E}\frac{|\eta_q(G/\sigma+Z;\chi^*)|^{q-2}}{(1+\chi^* q(q-1)|\eta_q(G/\sigma+Z;\chi^*)|^{q-2})^3}}_{J_1} \\ &\quad + q^2(1-q)\underbrace{\mathbb{E}\frac{\chi^*|\eta_q(G/\sigma+Z;\chi^*)|^{2q-4}}{(1+\chi^* q(q-1)|\eta_q(G/\sigma+Z;\chi^*)|^{q-2})^3}}_{J_2}. \end{aligned}$$

It is straightforward to see that

$$
\begin{aligned}
J_1 &\leq \mathbb{E}\frac{1}{|\eta_q(G/\sigma+Z;\chi^*)|^{2-q}+\chi^* q(q-1)}, \\
J_2 &\leq \mathbb{E}\frac{1/(q(q-1))}{|\eta_q(G/\sigma+Z;\chi^*)|^{2-q}+\chi^* q(q-1)}.
\end{aligned}
$$

We would like to prove $H_2 \to 0$ by showing

$$
\lim_{\sigma\to 0}\mathbb{E}\frac{1}{|\eta_q(G/\sigma+Z;\chi^*)|^{2-q}+\chi^* q(q-1)} = 0. \tag{2.84}
$$

Note that DCT may not be directly applied here, because the function inside the expectation cannot be easily bounded. Alternatively, we prove (2.84) by breaking the expectation into different parts and showing each part converges to zero. Let α_1 be a number that satisfies $\eta_q(\alpha_1;\chi^*) = (\chi^*)^{\frac{1}{2-q}}, \alpha_2 = (\chi^*)^{\frac{1}{2}}$, and α_3 a fixed positive constant that does not depend on σ. Denote the distribution of $|G|$ by $F(g)$. Note the following simple fact about α_1 according to Lemma 2.5.5 part (i):

$$
\alpha_1 = \eta_q(\alpha_1;\chi^*) + \chi^* q \eta_q^{q-1}(\alpha_1;\chi^*) = (q+1)(\chi^*)^{\frac{1}{2-q}}.
$$

So $\alpha_1 < \alpha_2 < \alpha_3$ when σ is small. Define the following three nested intervals:

$$
\mathcal{I}_i(x) \triangleq [-x-\alpha_i, -x+\alpha_i], \quad i = 1,2,3.
$$

With these definitions, we start the proof of (2.84). We have

$$
\begin{aligned}
&\mathbb{E}\frac{1}{|\eta_q(G/\sigma+Z;\chi^*)|^{2-q}+\chi^* q(q-1)} \\
=& \int_0^\infty \int_{z\in\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi^*)|^{2-q}+\chi^* q(q-1)}\phi(z)dzdF(g) + \\
& \int_0^\infty \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi^*)|^{2-q}+\chi^* q(q-1)}\phi(z)dzdF(g) + \\
& \int_0^\infty \int_{z\in\mathcal{I}_3(b/\sigma)\backslash\mathcal{I}_2(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi^*)|^{2-q}+\chi^* q(q-1)}\phi(z)dzdF(g) + \\
& \int_0^\infty \int_{\mathbb{R}\backslash\mathcal{I}_3(b/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi^*)|^{2-q}+\chi^* q(q-1)}\phi(z)dzdF(g) \\
\triangleq& \; G_1 + G_2 + G_3 + G_4,
\end{aligned}
$$

where $\phi(\cdot)$ is the density function of standard normal. We now bound each of the four integrals in (2.85) respectively.

$$\begin{aligned} G_1 &\leq \int_0^\infty \int_{z\in\mathcal{I}_1(g/\sigma)} \frac{1}{\chi^* q(q-1)}\phi(z)dzdF(g) \\ &\leq \frac{2\alpha_1\phi(0)}{\chi^* q(q-1)} \leq \frac{2(q+1)(\chi^*)^{\frac{1}{2-q}}\phi(0)}{\chi^* q(q-1)} \to 0, \text{ as } \sigma \to 0. \end{aligned}$$

The last step is due to the fact that $\chi^* \to 0$ by Lemma 2.5.30. For G_2 we have

$$\begin{aligned} G_2 &\leq \int_0^\infty \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi^*)|^{2-q}}\phi(z)dzdF(g) \\ &\overset{(b)}{\leq} \int_0^\infty \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{\chi^*}\phi(z)dzdF(g) \\ &\leq \int_0^{\sigma\log(1/\sigma)} \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{\chi^*}\phi(z)dzdF(g) + \\ &\quad \int_{\sigma\log(1/\sigma)}^\infty \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{\chi^*}\phi(z)dzdF(g) \\ &\leq \mathbb{P}(|G|\leq\sigma\log(1/\sigma))\frac{2\phi(0)\alpha_2}{\chi^*} + \frac{2\phi(\log(1/\sigma)-\alpha_2)\alpha_2}{\chi^*} \\ &= \mathbb{P}(|G|\leq\sigma\log(1/\sigma))\frac{2\phi(0)}{(\chi^*)^{1/2}} + \frac{2\phi(\log(1/\sigma)-\alpha_2)}{(\chi^*)^{1/2}} \\ &\overset{(c)}{\leq} O(1)\cdot\sigma^{2-q}\log(1/\sigma) + O(1)\cdot\frac{\phi((\log(1/\sigma))/2)}{\sigma^{q-1}} \to 0, \text{ as } \sigma\to 0. \end{aligned}$$

In the above derivations, (b) is because

$$|\eta_q(g/\sigma+z;\chi^*)| \geq \eta_q(\alpha_1;\chi^*) = (\chi^*)^{1/(2-q)} \text{ for } z \notin \mathcal{I}_1(g/\sigma).$$

To obtain (c), we have used the condition $\mathbb{P}(|G|\leq\sigma\log(1/\sigma)) = O(\sigma\log(1/\sigma))$ and the result $\chi^* = \Omega(\sigma^{2q-2})$ we proved at the beginning. Regarding G_3,

$$\begin{aligned} G_3 &\leq \int_0^\infty \int_{z\in\mathcal{I}_3(g/\sigma)\backslash\mathcal{I}_2(g/\sigma)} \frac{1}{|\eta_q(\alpha_2;\chi^*)|^{2-q}}\phi(z)dzdF(g) \\ &\leq \int_0^{\sigma\log 1/\sigma} \int_{z\in\mathcal{I}_3(g/\sigma)\backslash\mathcal{I}_2(g/\sigma)} \frac{1}{|\eta_q(\alpha_2;\chi^*)|^{2-q}}\phi(z)dzdF(g) \\ &\quad + \int_{\sigma\log 1/\sigma}^\infty \int_{z\in\mathcal{I}_3(g/\sigma)\backslash\mathcal{I}_2(g/\sigma)} \frac{1}{|\eta_q(\alpha_2;\chi^*)|^{2-q}}\phi(z)dzdF(g) \end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}(|G| \leq \sigma \log(1/\sigma)) \frac{2\phi(0)\alpha_3}{|\eta_q(\alpha_2;\chi^*)|^{2-q}} + \frac{2\phi(\log(1/\sigma)-\alpha_3)\alpha_3}{|\eta_q(\alpha_2;\chi^*)|^{2-q}} \\
&\overset{(d)}{\leq} O(1)\cdot \sigma^{q^2-3q+3}\log(1/\sigma) + O(1)\cdot \frac{\phi((\log(1/\sigma))/2)}{\sigma^{(q-1)(2-q)}} \to 0, \text{ as } \sigma \to 0.
\end{aligned}
$$

The calculations above are similar to those for G_2. In (d) we have used the following result:

$$
\lim_{\sigma\to 0} \frac{\eta_q(\alpha_2;\chi^*)}{(\chi^*)^{1/2}} = \lim_{\sigma\to 0} \frac{\eta_q((\chi^*)^{1/2};\chi^*)}{(\chi^*)^{1/2}} = \lim_{\sigma\to 0} \eta_q(1;(\chi^*)^{q/2}) = 1.
$$

Finally we can apply DCT to obtain

$$
\lim_{\sigma\to 0} G_4 = \mathbb{E} \lim_{\sigma\to 0} \frac{\mathbb{I}(|G/\sigma+Z| > \alpha_3)}{|\eta_q(G/\sigma+Z;\chi^*)|^{2-q} + \chi^* q(q-1)} = 0.
$$

We have finished the proof of $\lim_{\sigma\to 0} H_2 = 0$. This fact together with (2.81), (2.82) and (2.83) gives us

$$
\lim_{\sigma\to 0} \frac{\chi^*}{\sigma^{2q-2}} = \frac{-(1-\epsilon)H_1 - \epsilon H_2}{\epsilon H_3/\sigma^{2-2q}} = \frac{(1-\epsilon)\mathbb{E}|Z|^q}{\epsilon q \mathbb{E}|G|^{2q-2}}.
$$

□

2.5.7.3 Characterizing the convergence rate of $R_q(\chi_q^*(\sigma),\sigma)$

Having derived the convergence rate of $\chi_q^*(\sigma)$ in Chapter 2.5.7.2, we aim to obtain the convergence rate for $R_q(\chi_q^*(\sigma),\sigma)$ in this section. Towards that goal, we first present a useful lemma.

Lemma 2.5.32. *Suppose $\mathbb{P}(|G| \leq t) = O(t)$ (as $t \to 0$) and $\mathbb{E}|G|^2 < \infty$. If $\chi(\sigma) = C\sigma^{2q-2}$ with a fixed number $C > 0$, then for $1 < q < 2$ we have*

$$
\lim_{\sigma\to 0} \frac{R_q(\chi(\sigma),\sigma) - 1}{\sigma^{2q-2}} = -2C(1-\epsilon)q\mathbb{E}|Z|^q + \epsilon C^2 q^2 \mathbb{E}|G|^{2q-2}. \tag{2.85}
$$

Moreover, if $\chi(\sigma) = o(\sigma^{2q-2})$ then

$$
\lim_{\sigma\to 0} \frac{R_q(\chi(\sigma),\sigma) - 1}{\sigma^{2q-2}} = 0. \tag{2.86}
$$

Proof. We first focus on the case $\chi(\sigma) = C\sigma^{2q-2}$. According to (2.80),

$$\begin{aligned} R_q(\chi,\sigma) - 1 &= \underbrace{(1-\epsilon)\mathbb{E}(\eta_q^2(Z;\chi) - Z^2)}_{R_1} + \underbrace{\epsilon\chi^2 q^2 \mathbb{E}|\eta_q(G/\sigma + Z;\chi)|^{2q-2}}_{R_2} \\ & \quad \underbrace{-2\epsilon\chi q(q-1)\mathbb{E}\frac{|\eta_q(G/\sigma+Z;\chi)|^{q-2}}{1+\chi q(q-1)|\eta_q(G/\sigma+Z;\chi)|^{q-2}}}_{R_3}. \end{aligned}$$

Now we calculate the limit of each of the terms individually. We have

$$\begin{aligned} R_1 &= (1-\epsilon)\mathbb{E}(\eta_q(Z;\chi)+Z)(\eta_q(Z;\chi)-Z) \\ &\overset{(a)}{=} -(1-\epsilon)\mathbb{E}(\eta_q(Z;\chi)+Z)(\chi q|\eta_q(Z;\chi)|^{q-1}\text{sign}(Z)) \\ &= -(1-\epsilon)\chi q(\mathbb{E}|\eta_q(Z;\chi)|^q + \mathbb{E}|Z||\eta_q(Z;\chi)|^{q-1}), \end{aligned}$$

where (a) is due to Lemma 2.5.5 part (i). Hence we obtain

$$\begin{aligned} \lim_{\sigma\to 0}\frac{R_1}{\sigma^{2q-2}} &= -C(1-\epsilon)q\lim_{\sigma\to 0}(\mathbb{E}|\eta_q(Z;\chi)|^q + \mathbb{E}|Z||\eta_q(Z;\chi)|^{q-1}) \quad (2.87) \\ &= -2C(1-\epsilon)q\mathbb{E}|Z|^q. \end{aligned}$$

The last equality is by Dominated Convergence Theorem (DCT). For R_2,

$$\begin{aligned} \lim_{\sigma\to 0}\frac{R_2}{\sigma^{2q-2}} &= \lim_{\sigma\to 0}\frac{\epsilon\chi^2 q^2\mathbb{E}|\eta_q(G/\sigma+Z;\chi)|^{2q-2}}{\sigma^{2q-2}} \quad (2.88) \\ &= \epsilon C^2 q^2 \lim_{\sigma\to 0}\mathbb{E}|\eta_q(G+\sigma Z;\chi\sigma^{2-q})|^{2q-2} \\ &= \epsilon C^2 q^2 \mathbb{E}|G|^{2q-2}. \end{aligned}$$

Regarding the term R_3, we would like to show that if $\chi(\sigma) = O(\sigma^{2q-2})$, then

$$\lim_{\sigma\to 0}\frac{\chi}{\sigma^{2q-2}}\mathbb{E}\frac{1}{|\eta_q(G/\sigma+Z;\chi)|^{2-q}+\chi q(q-1)} = 0. \quad (2.89)$$

Define $\alpha_1 = 1$ if $1 < q < 3/2$ and $\alpha_1 = \sigma^{2q-3+c}$ if $3/2 \le q < 2$, where $c > 0$ is a sufficiently small constant that we will specify later. Let $F(g)$ be the distribution function of $|G|$ and $\alpha_2 > 1$ a fixed constant. So $\alpha_1 < \alpha_2$ for small σ. Define the following two nested intervals

$$\mathcal{I}_i(x) \triangleq [-x-\alpha_i, -x+\alpha_i], \quad i = 1, 2.$$

Then the expression in (2.89) can be written as

$$\begin{aligned}
& \frac{\chi}{\sigma^{2q-2}}\mathbb{E}\frac{1}{|\eta_q(G/\sigma+Z;\chi)|^{2-q}+\chi q(q-1)} \\
= & \frac{\chi}{\sigma^{2q-2}}\int_0^\infty \int_{z\in\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(g)+ \\
& \frac{\chi}{\sigma^{2q-2}}\int_0^\infty \int_{z\in\mathcal{I}_2(g/\sigma)\backslash\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(g)+ \\
& \frac{\chi}{\sigma^{2q-2}}\int_0^\infty \int_{\mathbb{R}\backslash\mathcal{I}_2(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(g) \\
\triangleq & G_1+G_2+G_3.
\end{aligned}$$

We will bound each of the three integrals above. The idea is similar as the one presented in the proof of Lemma 2.5.31. For the first integral,

$$\begin{aligned}
G_1 & \leq \frac{\chi}{\sigma^{2q-2}}\int_0^\infty \int_{-g/\sigma-\alpha_1}^{-g/\sigma+\alpha_1} \frac{1}{\chi q(q-1)}\phi(z)dzdF(g) \\
& \leq \frac{\chi}{\sigma^{2q-2}}\int_0^{\sigma\log 1/\sigma} \int_{-g/\sigma-\alpha_1}^{-g/\sigma+\alpha_1} \frac{1}{\chi q(q-1)}\phi(z)dzdF(g) \\
& \quad +\frac{\chi}{\sigma^{2q-2}}\int_{\sigma\log 1/\sigma}^\infty \int_{-g/\sigma-\alpha_1}^{-g/\sigma+\alpha_1} \frac{1}{\chi q(q-1)}\phi(z)dzdF(g) \\
& \overset{(b)}{\leq} \mathbb{P}(|G|\leq\sigma\log 1/\sigma)\frac{2\alpha_1\phi(0)}{q(q-1)\sigma^{2q-2}}+\frac{2\alpha_1\phi(\log 1/\sigma-\alpha_1)}{q(q-1)\sigma^{2q-2}} \\
& \leq O(1)\cdot\sigma^c\log(1/\sigma)\frac{\alpha_1}{\sigma^{2q-3+c}}+O(1)\cdot\frac{\alpha_1\phi((\log 1/\sigma)/2)}{\sigma^{2q-2}} \\
& \overset{(c)}{\to} 0, \text{ as } \sigma\to 0.
\end{aligned}$$

To obtain (b), we have used the following inequalities when σ is small:

$$\begin{aligned}
\int_{-g/\sigma-\alpha_1}^{-g/\sigma+\alpha_1}\phi(z)dz & \leq 2\phi(0)\alpha_1, \quad \text{for } g\leq\sigma\log(1/\sigma), \\
\int_{-g/\sigma-\alpha_1}^{-g/\sigma+\alpha_1}\phi(z)dz & \leq 2\alpha_1\phi(\log(1/\sigma)-\alpha_1), \quad \text{for } g>\sigma\log(1/\sigma).
\end{aligned}$$

The limit (c) holds due to the choice of α_1. Regarding the second term G_2,

$$\begin{aligned}
G_2 &\leq \frac{\chi}{\sigma^{2q-2}} \int_0^{\infty} \int_{z \in \mathcal{I}_2(g/\sigma)\setminus\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}} \phi(z) dz dF(g) \\
&= \frac{\chi}{\sigma^{2q-2}} \int_0^{\sigma \log 1/\sigma} \int_{z \in \mathcal{I}_2(g/\sigma)\setminus\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}} \phi(z) dz dF(g) \\
&\quad + \frac{\chi}{\sigma^{2q-2}} \int_{\sigma \log 1/\sigma}^{\infty} \int_{z \in \mathcal{I}_2(g/\sigma)\setminus\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(g/\sigma+z;\chi)|^{2-q}} \phi(z) dz dF(g) \\
&\overset{(d)}{\leq} \frac{\chi}{\sigma^{2q-2}} \int_0^{\sigma \log 1/\sigma} \int_{z \in \mathcal{I}_2(g/\sigma)\setminus\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(\alpha_1;\chi)|^{2-q}} \phi(z) dz dF(g) \\
&\quad + \frac{\chi}{\sigma^{2q-2}} \int_{\sigma \log 1/\sigma}^{\infty} \int_{z \in \mathcal{I}_2(g/\sigma)\setminus\mathcal{I}_1(g/\sigma)} \frac{1}{|\eta_q(\alpha_1;\chi)|^{2-q}} \phi(z) dz dF(g) \\
&\overset{(e)}{\leq} \mathbb{P}(|G| \leq \sigma \log 1/\sigma) \frac{2\alpha_2 \phi(0)\chi}{\sigma^{2q-2}|\eta_q(\alpha_1;\chi)|^{2-q}} + \frac{2\alpha_2 \phi(\log 1/\sigma - \alpha_2)\chi}{\sigma^{2q-2}|\eta_q(\alpha_1;\chi)|^{2-q}} \\
&\overset{(f)}{\leq} O(1) \cdot \sigma^c \log(1/\sigma) \frac{1}{\alpha_1^{2-q}\sigma^{c-1}} + O(1) \cdot \frac{\phi((\log 1/\sigma)/2)}{\alpha_1^{2-q}} \overset{(g)}{\to} 0, \text{ as } \sigma \to 0,
\end{aligned}$$

where (d) is due to the fact $|\eta_q(g/\sigma+z;\chi)| \geq \eta_q(\alpha_1;\chi)$ for $z \notin \mathcal{I}_1(g/\sigma)$; The argument for (e) is similar to that for (b); (f) holds based on two facts:

1. $\lim_{\sigma\to 0} \frac{\eta_q(\alpha_1;\chi)}{\alpha_1} = \lim_{\sigma \to 0} \eta_q(1; \alpha_1^{q-2}\chi) = 1$, since $\alpha_1^{q-2}\chi \to 0$. This is obvious for the case $\alpha_1 = 1$. When $\alpha_1 = \sigma^{2q-3+c}$, we have $\alpha_1^{q-2}\chi = O(1) \cdot \sigma^{2q^2+(c-5)q+4-2c}$ and $2q^2 + (c-5)q + 4 - 2c > 0$ if c is chosen small enough.

2. $\mathbb{P}(|G| \leq \sigma \log 1/\sigma) = O(\sigma \log 1/\sigma)$. This is one of the conditions.

And finally (g) works as follows: it is clear that $\sigma^c \log(1/\sigma) \frac{1}{\alpha_1^{2-q}\sigma^{c-1}}$ goes to zero when $\alpha_1 = 1$; when $\alpha_1 = \sigma^{2q-3+c}$, we can sufficiently small c such that $\alpha_1^{q-2}\sigma^{1-c} = \sigma^{2q^2+(c-7)q+7-3c} = o(1)$. For the third integral G_3, we are able to invoke DCT to obtain

$$\lim_{\sigma \to 0} G_3 = O(1) \cdot \lim_{\sigma \to 0} \mathbb{E} \frac{\mathbb{I}(|G/\sigma + Z| > \alpha_2)}{|\eta_q(G/\sigma + Z;\chi)|^{2-q} + \chi q(q-1)} = 0.$$

Note that DCT works because for small σ

$$\begin{aligned}
&\frac{\mathbb{I}(|G/\sigma + Z| > \alpha_2)}{|\eta_q(G/\sigma + Z;\chi)|^{2-q} + \chi q(q-1)} \\
\leq\ & \frac{\mathbb{I}(|G/\sigma + Z| > \alpha_2)}{|\eta_q(G/\sigma + Z;\chi)|^{2-q}} \leq \frac{1}{|\eta_q(\alpha_2;\chi)|^{2-q}} \leq \frac{1}{|\eta_q(\alpha_2;1)|^{2-q}}.
\end{aligned}$$

We hence have showned that

$$\lim_{\sigma \to 0} \frac{R_3}{\sigma^{2q-2}} = 0. \tag{2.90}$$

Combining the results (2.87), (2.88), and (2.90) establishes (2.85).

To prove (2.86), first note that (2.90) has been derived in the general setting $\chi(\sigma) = O(\sigma^{2q-2})$. Moreover, we can use similar arguments to show for $\chi(\sigma) = o(\sigma^{2q-2})$,

$$\lim_{\sigma \to 0} \frac{R_1}{\sigma^{2q-2}} = \lim_{\sigma \to 0} \frac{R_2}{\sigma^{2q-2}} = 0.$$

This completes the proof. □

We are in position to derive the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$.

Lemma 2.5.33. *Suppose* $\mathbb{P}(|G| \leq t) = O(t)$ *(as* $t \to 0$*) and* $\mathbb{E}|G|^2 < \infty$*. Then for* $q \in (1, 2)$ *we have*

$$\lim_{\sigma \to 0} \frac{R_q(\chi_q^*(\sigma), \sigma) - 1}{\sigma^{2q-2}} = -\frac{(1-\epsilon)^2 (\mathbb{E}|Z|^q)^2}{\epsilon \mathbb{E}|G|^{2q-2}}.$$

Proof. According to Lemma 2.5.31, choosing $C = \frac{(1-\epsilon)\mathbb{E}|Z|^q}{\epsilon q \mathbb{E}|G|^{2q-2}}$ in Lemma 2.5.32 finishes the proof. □

2.5.7.4 Deriving the expansion of AMSE$(\lambda_{*,q}, q, \sigma_w)$

According to Corollary 2.2.4 we know

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \bar{\sigma}^2 \cdot R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}), \tag{2.91}$$

where $\bar{\sigma}$ satisfies the following equation:

$$\bar{\sigma}^2 = \sigma_w^2 + \frac{\bar{\sigma}^2}{\delta} R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}). \tag{2.92}$$

Since $\chi_q^*(\bar{\sigma})$ minimizes $R_q(\chi, \bar{\sigma})$ we obtain

$$R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) \leq R_q(0, \bar{\sigma}) = 1.$$

Therefore, under the condition $\delta > 1$,

$$\bar{\sigma}^2 \leq \sigma_w^2 + \frac{1}{\delta}\bar{\sigma}^2 \Rightarrow \bar{\sigma}^2 \leq \frac{\delta}{\delta - 1}\sigma_w^2, \tag{2.93}$$

which implies that $\bar{\sigma} \to 0$ as $\sigma_w \to 0$. Accordingly, we combine Equation (2.92) with the fact $R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) \to 1$ as $\bar{\sigma} \to 0$ from Lemma 2.5.33 to conclude

$$\lim_{\sigma_w \to 0} \frac{\sigma_w^2}{\bar{\sigma}^2} = \lim_{\bar{\sigma} \to 0} \frac{\sigma_w^2}{\bar{\sigma}^2} = \lim_{\bar{\sigma} \to 0} \Big(1 - \frac{1}{\delta} R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma})\Big) = \frac{\delta - 1}{\delta}. \tag{2.94}$$

We now derive the expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ presented in (2.11). From (2.91) and (2.92) we can compute that

$$\begin{aligned} \frac{\text{AMSE}(\lambda_{*,q}, q, \sigma_w) - \frac{\sigma_w^2}{1-1/\delta}}{\sigma_w^{2q}} &= \frac{\bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - \frac{1}{1-1/\delta} \cdot (\bar{\sigma}^2 - \frac{\bar{\sigma}^2}{\delta} R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}))}{\sigma_w^{2q}} \\ = \quad \frac{\bar{\sigma}^2 \delta (R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1)}{\sigma_w^{2q}(\delta - 1)} &= \frac{\bar{\sigma}^{2q}}{\sigma_w^{2q}} \cdot \frac{\delta}{\delta - 1} \cdot \frac{R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1}{\bar{\sigma}^{2q-2}}. \end{aligned}$$

Letting $\sigma_w \to 0$ on both sides of the above equation and using the results from (2.94) and Lemma 2.5.33 completes the proof.

2.5.8 Proof of Theorem 2.2.6

From (2.91) and (2.92) we see that

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \delta(\bar{\sigma}^2 - \sigma_w^2).$$

Hence Theorem 2.2.6 can be proved by showing

$$\lim_{\sigma_w \to 0} \bar{\sigma} > 0.$$

For that purpose we first prove the following under the condition $\mathbb{E}|G|^2 < \infty$.

$$\lim_{\sigma \to 0} R_q(\chi_q^*(\sigma), \sigma) = 1. \tag{2.95}$$

When $q = 2$, $R_q(\chi_q^*(\sigma); \sigma)$ admits a nice explicit expression and can be easily shown to converge to 1. For $1 < q < 2$, since $\chi_q^*(\sigma)$ is the minimizer of $R_q(\chi, \sigma)$ we know

$$R_q(\chi_q^*(\sigma), \sigma) \leq R_q(0, \sigma) = 1,$$

hence $\limsup_{\sigma \to 0} R_q(\chi_q^*(\sigma), \sigma) \leq 1$. On the other hand, (2.80) gives us

$$\begin{aligned} R_q(\chi_q^*(\sigma), \sigma) &\geq (1-\epsilon)\mathbb{E}(\eta_q(Z;\chi^*))^2 - \epsilon \\ &\quad + 2\epsilon \mathbb{E}\left(\frac{1}{1+\chi^* q(q-1)|\eta_q(G/\sigma + Z;\chi^*)|^{q-2}} \right) \end{aligned}$$

where we have used χ^* to denote $\chi_q^*(\sigma)$ for simplicity. Based on Lemmas 2.5.29 and 2.5.30, we can apply Fatou's lemma to the above inequality to obtain $\liminf_{\sigma \to 0} R_q(\chi_q^*(\sigma), \sigma) \geq 1$.

Next it is clear that

$$\lim_{\sigma \to \infty} R_q(\chi_q^*(\sigma), \sigma) \leq \lim_{\sigma \to \infty} \lim_{\chi \to \infty} R_q(\chi, \sigma) = 0. \tag{2.96}$$

We now consider an arbitrary convergent sequence $\bar{\sigma}_n \to \sigma^*$. We claim $\sigma^* \neq 0$. Otherwise Equation (2.92) tells us

$$R_q(\chi_q^*(\bar{\sigma}_n), \bar{\sigma}_n) < \delta < 1,$$

and letting $n \to \infty$ above contradicts (2.95). Now that $\sigma^* > 0$ we can take $n \to \infty$ in (2.92) to obtain

$$R_q(\chi_q^*(\sigma^*), \sigma^*) = \delta < 1.$$

According to Lemma 2.5.13, it is not hard to confirm $R_q(\chi_q^*(\sigma), \sigma)$ is a strictly decreasing and continuous function of σ. Results (2.95) and (2.96) then imply that σ^* is the unique solution to $R_q(\chi_q^*(\sigma), \sigma) = \delta$. Since this is true for any sequence, we have proved $\lim_{\sigma_w \to 0} \bar{\sigma}$ exists and larger than zero.

2.5.9 Proof of Theorem 2.2.8

2.5.9.1 Roadmap

Theorem 2.2.8 differs from Theorem 2.2.7 in that the order of the second dominant term of $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$ becomes polynomial (ignore the logarithm term) when the distribution of G has mass around zero. However, the proof outline remains the same.

We hence stick to the same notations used in the proof of Theorem 2.2.7. In particular, $R(\chi, \sigma), \chi^*(\sigma)$ represent $R_q(\chi, \sigma), \chi_q^*(\sigma)$ with $q = 1$, respectively. We characterize the convergence rate of $\chi^*(\sigma)$ in Chapter 2.5.9.2, and bound the convergence rate of $R(\chi^*(\sigma), \sigma)$ in Chapter 2.5.9.3. After we characterize $R(\chi^*(\sigma), \sigma)$, the rest of the proof is similar to that in Chapter 2.5.2.4. We therefore do not repeat it here.

2.5.9.2 Bounding the convergence rate of $\chi^*(\sigma)$

Lemma 2.5.34. *Suppose $\mathbb{P}(|G| \leq t) = \Theta(t^\ell)$ with $\ell > 0$ (as $t \to 0$) and $\mathbb{E}|G|^2 < \infty$, then for sufficiently small σ*

$$\alpha_m \sigma^\ell \leq \chi^*(\sigma) - \chi^{**} \leq \beta_m \sigma^\ell \cdot \left(\sqrt{\underbrace{\log \log \ldots \log}_{m \text{ times}} \left(\frac{1}{\sigma} \right)} \right)^\ell,$$

where $m > 0$ is an arbitrary integer number, $\alpha_m, \beta_m > 0$ are two constants depending on m, and χ^{**} is the unique minimizer of $(1-\epsilon)\mathbb{E}(\eta_1(Z;\chi))^2 + \epsilon(1+\chi^2)$ over $[0, \infty)$.

Proof. According to Lemma 2.5.1, $\chi^*(\sigma) \to \chi^{**}$ as $\sigma \to 0$. To characterize the convergence rate, we follow the same line of proof that we presented for Lemma 2.5.2 and adopt the same notations. For simplicity we do not detail out the entire proof and instead highlight the differences. The key difference is that neither e_1 or e_2 are exponentially small in the current setting. We now start by bounding e_2. Let $F(g)$ be the distribution function of $|G|$ and define

$$\log_m(a) \triangleq \underbrace{\log \log \ldots \log}_{m \text{ times}}(a).$$

Given an integer $m > 0$ and a constant $c > 0$, we then have

$$\begin{aligned}
&\mathbb{E}\phi(\chi^* - |G|/\sigma) \\
&= \sum_{i=1}^{m-1} \int_{c\sigma(\log_{m-i+1}(1/\sigma))^{1/2}}^{c\sigma(\log_{m-i}(1/\sigma))^{1/2}} \phi(\chi^* - g/\sigma)dF(g) + \\
&\quad \int_0^{c\sigma(\log_m(1/\sigma))^{1/2}} \phi(\chi^* - g/\sigma)dF(g) + \int_{c\sigma(\log(1/\sigma))^{1/2}}^{\infty} \phi(\chi^* - g/\sigma)dF(g) \\
&\leq \sum_{i=1}^{m-1} \phi(c(\log_{m-i+1}(1/\sigma))^{1/2} - \chi^*) \cdot \mathbb{P}(|G| \leq c\sigma(\log_{m-i}(1/\sigma))^{1/2}) + \\
&\qquad \phi(0) \cdot \mathbb{P}(|G| \leq c\sigma(\log_m(1/\sigma))^{1/2}) + \phi(c(\log(1/\sigma))^{1/2} - \chi^*). \qquad (2.97)
\end{aligned}$$

The condition $\mathbb{P}(|G| \leq t) = \Theta(t^\ell)$ leads to

$$\begin{aligned}
&\sum_{i=1}^{m-1} \phi(c(\log_{m-i+1}(1/\sigma))^{1/2} - \chi^*) \cdot \mathbb{P}(|G| \leq c\sigma(\log_{m-i}(1/\sigma))^{1/2}) \\
&\quad \leq \frac{e^{(\chi^*)^2/2}}{\sqrt{2\pi}} \sum_{i=1}^{m-1} \mathrm{e}^{-(c^2 \log_{m-i+1}(1/\sigma))/4} \cdot \Theta(\sigma^\ell (\log_{m-i}(1/\sigma))^{\ell/2}) \\
&\quad = O(1) \cdot \sum_{i=1}^{m-1} \sigma^\ell (\log_{m-i}(1/\sigma))^{\ell/2 - c^2/4},
\end{aligned}$$

where we have used the simple inequality $e^{-(a-b)^2/2} \leq e^{-b^2/4} \cdot e^{a^2/2}$. It is also clear that

$$\phi(c(\log(1/\sigma))^{1/2} - \chi^*) \leq \frac{1}{\sqrt{2\pi}} e^{(\chi^*)^2/2} \cdot \sigma^{c^2/4}.$$

Therefore, by choosing a sufficiently large c we can conclude that the dominant term in (2.97) is $\phi(0)\mathbb{P}(|G| \leq c\sigma(\log_m(1/\sigma))^{1/2}) = \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2})$. Furthermore, choosing a fixed constant $C > 0$ we have the following lower bound

$$\mathbb{E}\phi(\chi^* + |G|/\sigma) \geq \int_0^{C\sigma} \phi(\chi^* + g/\sigma)dF(g) \geq \phi(C + \chi^*) \cdot \mathbb{P}(|G| \leq C\sigma) = \Theta(\sigma^\ell).$$

Because

$$\mathbb{E}\phi(\chi^* + |G|/\sigma) \leq \mathbb{E}\phi(\chi^* \pm G/\sigma) \leq \mathbb{E}\phi(\chi^* - |G|/\sigma),$$

We are able to derive

$$\Theta(\sigma^\ell) \leq \mathbb{E}\phi(\chi^* \pm G/\sigma) \leq \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2}).$$

As a result we obtain the bound for e_2:

$$\Theta(\sigma^\ell) \leq e_2 \leq \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2}) \tag{2.98}$$

To bound e_1, recall that

$$e_1 = -\epsilon\mathbb{E}\int_{-G/\sigma-\chi^*}^{-G/\sigma+\chi^*} \phi(z)dz = -2\epsilon\chi^*\mathbb{E}\phi(a\chi^* - G/\sigma),$$

where $|a| \leq 1$ depends on G. We can find two positive constants $C_1, C_2 > 0$ such that for small σ

$$C_1\mathbb{E}\phi(\chi^* + |G|/\sigma) \leq \mathbb{E}\phi(a\chi^* - G/\sigma) \leq C_2\mathbb{E}\phi(\chi^* - |G|/\sigma).$$

Hence we have

$$\Theta(\sigma^\ell) \leq -e_1 \leq \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2}). \tag{2.99}$$

Based on the results from (2.98) and (2.99) and Equation (2.24), we can use similar arguments as in the proof of Lemma 2.5.2 to conclude

$$\Theta(\sigma^\ell) \leq \chi^*(\sigma) - \chi^{**} \leq \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2}).$$

□

2.5.9.3 Bounding the convergence rate of $R(\chi^*(\sigma), \sigma)$

Lemma 2.5.35. *Suppose $\mathbb{P}(|G| \leq t) = \Theta(t^\ell)$ with $\ell > 0$ (as $t \to 0$) and $\mathbb{E}|G|^2 < \infty$, then for sufficiently small σ*

$$-\beta_m\sigma^\ell \cdot \left(\sqrt{\underbrace{\log\log\ldots\log}_{m\ \text{times}}\left(\frac{1}{\sigma}\right)}\right)^\ell \leq R(\chi^*(\sigma), \sigma) - M_1(\epsilon) \leq -\alpha_m\sigma^\ell,$$

where $m > 0$ is an arbitrary integer number and $\alpha_m, \beta_m > 0$ are two constants depending on m.

Proof. We recall the two quantities:

$$\begin{aligned} M_1(\epsilon) &= (1-\epsilon)\mathbb{E}(\eta_1(Z;\chi^{**}))^2 + \epsilon(1+(\chi^{**})^2) \\ R(\chi^*(\sigma),\sigma) &= (1-\epsilon)\mathbb{E}(\eta_1(Z;\chi^*))^2 + \epsilon(1+\mathbb{E}(\eta_1(G/\sigma+Z;\chi^*)-G/\sigma-Z)^2) \\ &\quad +2\epsilon\mathbb{E}Z(\eta_1(G/\sigma+Z;\chi^*)-G/\sigma-Z). \end{aligned}$$

Since $\chi^*(\sigma)$ is the minimizer of $R(\chi,\sigma)$,

$$\begin{aligned} & R(\chi^*(\sigma),\sigma) - M_1(\epsilon) \le R(\chi^{**},\sigma) - M_1(\epsilon) \\ = \;& \epsilon[\mathbb{E}(\eta_1(G/\sigma+Z;\chi^{**})-G/\sigma-Z)^2 - (\chi^{**})^2] + \\ & 2\epsilon\mathbb{E}Z(\eta_1(G/\sigma+Z;\chi^{**})-G/\sigma-Z) \\ \overset{(a)}{\le} \;& -2\epsilon\mathbb{E}\mathbb{I}(|G/\sigma+Z|\le\chi^{**}) \overset{(b)}{\le} -\Theta(\sigma^\ell). \end{aligned} \tag{2.100}$$

To obtain (a), we have used Lemma 2.5.12 and the fact $|\eta_1(u;\chi)-u|\le\chi$. (b) is due to the similar arguments for bounding e_1 in Lemma 2.5.34. To derive the lower bound for $R(\chi^*(\sigma),\sigma)-M_1(\epsilon)$, we can follow the same reasoning steps as in the proof of Lemma 2.5.3 and utilize the bound we derived for $|\chi^*(\sigma)-\chi^{**}|$ in Lemma 2.5.34. We will obtain

$$|R(\chi^*(\sigma),\sigma)-M_1(\epsilon)| \le \Theta(\sigma^\ell(\log_m(1/\sigma))^{\ell/2}). \tag{2.101}$$

Putting (2.100) and (2.101) together completes the proof. □

2.5.10 Proof of Theorem 2.2.9

The proof of Theorem 2.2.9 is essentially the same as that of Theorem 2.2.6. We do not repeat the details. Note that the key argument $\lim_{\sigma\to 0} R(\chi^*(\sigma),\sigma) = M_1(\epsilon)$ has been shown in Lemma 2.5.1.

2.5.11 Proof of Theorem 2.2.10

Similarly as in the proof of Theorems 2.2.5, we would like to derive the convergence rate of $R_q(\chi_q^*(\sigma),\sigma)$. We first characterize the convergence rate of $\chi_q^*(\sigma)$.

Lemma 2.5.36. *Suppose* $\mathbb{E}|G|^2 < \infty$*, then for* $0 < q < 1$*,* $\chi_q^*(\sigma) \to \infty$ *and* $\chi_q^*(\sigma)\sigma^{2-q} \to 0$*, as* $\sigma \to 0$*.*

Proof. If $\chi_q^*(\sigma)\sigma^{2-q} \nrightarrow 0$, then there exists a sequence $\sigma_k \to 0$ and a constant $c > 0$ such that $\chi_q^*(\sigma_k)\sigma_k^{2-q} \geq c$, for all k. Choose a convergent subsequence $\{\sigma_{k_n}\}$ and denote $\lim_{k_n\to\infty} \chi_q^*(\sigma_{k_n})\sigma_{k_n}^{2-q} = \alpha \geq c$ (note α can be $+\infty$). Fatou's lemma gives

$$\begin{aligned}
& \liminf_{k_n\to\infty} \mathbb{E}(\eta_q(B+\sigma_{k_n}Z;\sigma_{k_n}^{2-q}\chi_q^*(\sigma_{k_n}))-B)^2 \\
\geq\ & \mathbb{E}\liminf_{k_n\to\infty}(\eta_q(B+\sigma_{k_n}Z;\sigma_{k_n}^{2-q}\chi_q^*(\sigma_{k_n}))-B)^2 \\
\geq\ & \mathbb{E}\min((\eta_q(B;\alpha)-B)^2, B^2) > 0
\end{aligned}$$

Hence, we have

$$\begin{aligned}
& \liminf_{k_n\to\infty} \mathbb{E}(\eta_q(B/\sigma_{k_n}+Z;\chi_q^*(\sigma_{k_n}))-B/\sigma_{k_n})^2 \\
= & \lim_{k_n\to\infty} \frac{1}{\sigma_{k_n}^2} \liminf_{k_n\to\infty} \mathbb{E}(\eta_q(B+\sigma_{k_n}Z;\sigma_{k_n}^{2-q}\chi_q^*(\sigma_{k_n}))-B)^2 = +\infty,
\end{aligned}$$

which implies $\liminf_{k_n\to\infty} R_q(\chi_q^*(\sigma_{k_n}),\sigma_{k_n}) = +\infty$. However, since $\chi_q^*(\sigma_{k_n})$ is the optimal thresholding value, it holds that $R_q(\chi_q^*(\sigma_{k_n}),\sigma_{k_n}) \leq R_q(0,\sigma_{k_n}) = 1$, for every k_n. This is a contradiction. Similarly, if $\chi_q^*(\sigma) \nrightarrow \infty$, there exists a sequence $\sigma_k \to 0$ and a finite constant $\alpha \geq 0$ such that $\chi_q^*(\sigma_k) \to \alpha$. We can apply Dominated Convergence Theorem (DCT) to obtain,

$$\lim_{k\to\infty} R_q(\chi_q^*(\sigma_k),\sigma_k) = (1-\epsilon)\mathbb{E}\eta_q^2(Z;\alpha) + \epsilon > \epsilon. \tag{2.102}$$

On the other hand, since $\chi_q^*(\sigma_k)$ is the optimal thresholding value, we have

$$\lim_{k\to\infty} R_q(\chi_q^*(\sigma_k),\sigma_k) \leq \lim_{k\to\infty} R_q(\beta,\sigma_k) = (1-\epsilon)\mathbb{E}\eta_q^2(Z;\beta) + \epsilon,$$

for any finite β. Letting $\beta \to \infty$ on both sides of the above inequality yields

$$\lim_{k\to\infty} R_q(\chi_q^*(\sigma_k),\sigma_k) \leq \epsilon,$$

which contradicts (2.102). □

Lemma 2.5.37. *Suppose $\mathbb{P}(|G| > \mu) = 1$ with μ being a fixed positive number and $\mathbb{E}|G|^2 < \infty$, then for every $0 < q < 1$,*

$$\lim_{\sigma \to 0} \frac{\sigma^{2-2q}}{(\chi_q^*(\sigma))^{\frac{2q-1}{2-q}} \phi(c_q (\chi_q^*(\sigma))^{\frac{1}{2-q}})} = \frac{(1-\epsilon) c_q (\eta_q^+(c_q; 1))^2}{\epsilon q^2 (2-q) \mathbb{E}|G|^{2q-2}},$$

where $\phi(\cdot)$ is the density function of a standard normal.

Proof. Let $F(g)$ denote the distribution function of $|G|$. We first decompose $R_q(\chi, \sigma)$.

$$\begin{aligned} R_q(\chi, \sigma) &= \underbrace{2(1-\epsilon) \int_{c_q \chi^{\frac{1}{2-q}}}^{\infty} \eta_q^2(z; \chi) \phi(z) dz}_{R_1} \\ &+ \underbrace{\epsilon \int_{\mu}^{\infty} \int_{-\frac{g}{\sigma} + c_q \chi^{\frac{1}{2-q}}}^{\infty} (\eta_q(\frac{g}{\sigma} + z; \chi) - \frac{g}{\sigma})^2 \phi(z) dz dF(g)}_{R_2} \\ &+ \underbrace{\epsilon \int_{\mu}^{\infty} \int_{\frac{g}{\sigma} + c_q \chi^{\frac{1}{2-q}}}^{\infty} (\eta_q(-\frac{g}{\sigma} + z; \chi) + \frac{g}{\sigma})^2 \phi(z) dz dF(g)}_{R_3} \\ &+ \underbrace{\epsilon \int_{\mu}^{\infty} \int_{-\frac{g}{\sigma} - c_q \chi^{\frac{1}{2-q}}}^{-g/\sigma + c_q \chi^{\frac{1}{2-q}}} \frac{g^2}{\sigma^2} \phi(z) dz dF(g)}_{R_4}. \end{aligned} \tag{2.103}$$

From the proof of Lemma 2.5.36, it is straightforward to see that $\chi_q^*(\sigma)$ is non-zero and finite. Since $\chi = \chi_q^*(\sigma)$ is the optimal thresholding value, we conclude that $\chi_q^*(\sigma)$ satisfies $\frac{\partial R_q(\chi_q^*(\sigma), \sigma)}{\partial \chi} = 0$. For notational simplicity, below we may write χ for $\chi_q^*(\sigma)$. Now we analyze the partial derivative of the four terms in (2.103) separately. For the first term,

$$\begin{aligned} \frac{\partial R_1}{\partial \chi} &= \frac{-2(1-\epsilon) c_q \chi^{\frac{q-1}{2-q}}}{2-q} (\eta_q^+(c_q \chi^{\frac{1}{2-q}}; \chi))^2 \phi(c_q \chi^{\frac{1}{2-q}}) \\ &\quad -4(1-\epsilon) q \int_{c_q \chi^{\frac{1}{2-q}}}^{\infty} \frac{\eta_q^q(z; \chi)}{1 + \chi q(q-1) \eta_q^{q-2}(z; \chi)} \phi(z) dz, \end{aligned} \tag{2.104}$$

where we have used Lemma 2.5.9 part (i). We now compare the order of the two terms on the right hand side of the above equality. According to Lemma 2.5.10, we can conclude that $1 + \chi q(q-1) \eta_q^{q-2}(z; \chi)$ is bounded away from zero, for $z \geq c_q \chi^{\frac{1}{2-q}}$.

Hence, combining with the fact $|\eta_q(z;\chi)| \leq |z|$, we obtain there exists a positive constant C such that

$$\begin{aligned}
&\Big|\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty} \frac{\eta_q^q(z;\chi)}{1+\chi q(q-1)\eta_q^{q-2}(z;\chi)}\phi(z)dz\Big| \leq C\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty} z^q\phi(z)dz \\
&\overset{(i)}{=} Cc_q^{q-1}\chi^{\frac{q-1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}}) + C\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}(q-1)z^{q-2}\phi(z)dz \\
&\leq Cc_q^{q-1}\chi^{\frac{q-1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}}) + C(q-1)c_q^{q-2}\chi^{-1}\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}\phi(z)dz \\
&\overset{(ii)}{\leq} O(\chi^{\frac{q-1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})) + O(\chi^{\frac{q-3}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})) = O(\chi^{\frac{q-1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})).
\end{aligned}$$

We have used integration by parts to obtain (i). Regarding (ii), we have used the tail approximation $\int_t^\infty \phi(z)dz \sim \frac{1}{t}\phi(t)$, as $t \to \infty$. Now we discuss the order of the first term in (2.104). Since $\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi) = \chi^{\frac{1}{2-q}}\eta_q^+(c_q;1)$ from Lemma 2.5.9 part (iv), we know the first term is of order $\chi^{\frac{q+1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})$. Hence we can conclude that

$$\lim_{\sigma\to 0}\frac{\partial R_1}{\partial\chi}/(\chi^{\frac{q+1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})) = \frac{-2(1-\epsilon)c_q(\eta_q^+(c_q;1))^2}{2-q}. \tag{2.105}$$

For the last term R_4, we can do the following calculations:

$$\begin{aligned}
\frac{\partial R_4}{\partial\chi} &= \mathbb{E}\Big[\frac{\epsilon|G|^2c_q\chi^{\frac{q-1}{2-q}}}{\sigma^2(2-q)}(\phi(-G/\sigma + c_q\chi^{\frac{1}{2-q}}) + \phi(-G/\sigma - c_q\chi^{\frac{1}{2-q}}))\Big] \\
&\leq \frac{2\epsilon c_q\chi^{\frac{q-1}{2-q}}}{\sigma^2(2-q)}\mathbb{E}[|G|^2\phi(c_q\chi^{\frac{1}{2-q}} - |G|/\sigma)] \leq \frac{2\epsilon c_q\tau^{\frac{q-1}{2-q}}}{\sigma^2(2-q)}\phi(c_q\chi^{\frac{1}{2-q}} - \mu/\sigma)\mathbb{E}|G|^2,
\end{aligned}$$

where the last inequality is due to the fact $|G|/\sigma \geq \mu/\sigma \gg c_q\chi^{\frac{1}{2-q}}$ from Lemma 2.5.36. Making use of $\mu/\sigma \gg c_q\chi^{\frac{1}{2-q}}$ again, it is straightforward to confirm that

$$\lim_{\sigma\to 0}\frac{\phi(-\mu/\sigma + c_q\chi^{\frac{1}{2-q}})}{\sigma^2\phi(c_q\chi^{\frac{1}{2-q}})} = 0.$$

Therefore we conclude

$$\lim_{\sigma\to 0}\frac{\partial R_4}{\partial\chi}/(\chi^{\frac{q+1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})) = 0. \tag{2.106}$$

We now discuss the calculation of $\frac{\partial R_2}{\partial \chi}$.

$$\begin{aligned}
\frac{\partial R_2}{\partial \chi} &= \frac{-\epsilon c_q}{2-q}\chi^{\frac{q-1}{2-q}}\int_{\mu}^{\infty}(\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi)-g/\sigma)^2\phi(-g/\sigma+c_q\chi^{\frac{1}{2-q}})dF(g)+ \\
&2\epsilon\int_{\mu}^{\infty}\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{\infty}(\eta_q(g/\sigma+z;\chi)-g/\sigma-z)\partial_2\eta_q(g/\sigma+z;\chi)\phi(z)dzdF(g) \\
&+2\epsilon\int_{\mu}^{\infty}\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{\infty}z\partial_2\eta_q(g/\sigma+z;\chi)\phi(z)dzdF(g) \\
&= \underbrace{\frac{-\epsilon c_q}{2-q}\chi^{\frac{q-1}{2-q}}\int_{\mu}^{\infty}(\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi)-g/\sigma)^2\phi(-g/\sigma+c_q\chi^{\frac{1}{2-q}})dF(g)}_{S_1}+ \\
&\underbrace{2\epsilon\chi q^2\int_{\mu}^{\infty}\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{\infty}\frac{\eta_q^{2q-2}(g/\sigma+z;\chi)}{1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi)}\phi(z)dzdF(g)}_{S_2}+ \\
&\underbrace{-2\epsilon q\int_{\mu}^{\infty}\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{\infty}\frac{\eta_q^{q-1}(g/\sigma+z;\chi)}{1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi)}z\phi(z)dzdF(g)}_{S_3},
\end{aligned}$$

where we have used Lemma 2.5.9 part (i)(iii) in the above derivations. We then analyze the above three terms separately. For S_3, integration by parts combined with Lemma 2.5.9 part (i) gives

$$\begin{aligned}
S_3 &= \underbrace{\frac{-2\epsilon q(\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi))^{q-1}}{1+\chi q(q-1)(\eta_q^+(c_q\chi^{\frac{1}{2-q}};\chi))^{q-2}}\int_{\mu}^{\infty}\phi(-\frac{g}{\sigma}+c_q\chi^{\frac{1}{2-q}})dF(g)}_{T_1}+ \\
&\underbrace{-2\epsilon q(q-1)\int_{\mu}^{\infty}\int_{-\frac{g}{\sigma}+c_q\chi^{\frac{1}{2-q}}}^{\infty}\frac{\eta_q^{q-2}(g/\sigma+z;\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi))^2}\phi(z)dzdF(g)}_{T_2}+ \\
&\underbrace{2\epsilon q^2(q-1)(q-2)\chi\int_{\mu}^{\infty}\int_{-\frac{g}{\sigma}+c_q\chi^{\frac{1}{2-q}}}^{\infty}\frac{\eta_q^{2q-4}(g/\sigma+z;\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi))^3}\phi(z)dzdF(g)}_{T_3}.
\end{aligned}$$

Choosing a positive constant $0<v<\mu$, we write

$$\begin{aligned}
\frac{T_2}{\sigma^{2-q}} &= -2\epsilon q(q-1)\int_{\mu}^{\infty}\int_{-\frac{g}{\sigma}+c_q\chi^{\frac{1}{2-q}}}^{-\frac{g}{\sigma}+\frac{v}{\sigma}}\frac{\sigma^{q-2}\eta_q^{q-2}(g/\sigma+z;\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi))^2}\phi(z)dzdF(g) \\
&-2\epsilon q(q-1)\mathbb{E}\left[\frac{\mathbb{1}(Z+|G|/\sigma>v/\sigma)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(|G|/\sigma+Z;\chi))^2}\right]. \qquad (2.107)
\end{aligned}$$

It is clear that when $u > c_q\chi^{\frac{1}{2-q}}$, there exists a positive constant C_0 such that $1 + \chi q(q-1)\eta_q^{q-2}(u;\chi) > C_0 > 0$. Also since $\eta_q(u;\chi)$ is a non-decreasing function of $u > 0$, we can obtain

$$\begin{aligned}&\left|\frac{\mathbb{1}(Z+|G|/\sigma > v/\sigma)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(|G|/\sigma+Z;\chi))^2}\right|\\ &\leq (C_0^2)^{-1}\eta_q^{q-2}(v;\sigma^{2-q}\chi) \leq C_0^{-2}\eta_q^{q-2}(v;1),\end{aligned}$$

for sufficiently small σ. Because $\sigma^{2-q}\chi \to 0$, as $\sigma \to 0$ from Lemma 2.5.36, we get

$$\begin{aligned}&\lim_{\sigma\to 0}\frac{\mathbb{1}(Z+|G|/\sigma > v/\sigma)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(|G|/\sigma+Z;\chi))^2}\\ &= \lim_{\sigma\to 0}\frac{\mathbb{1}(\sigma Z+|G| > v)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\sigma^{2-q}\chi q(q-1)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi))^2} = |G|^{q-2}.\end{aligned}$$

We can then use Dominated Convergence Theorem (DCT) to conclude

$$\lim_{\sigma\to 0}\mathbb{E}\frac{\mathbb{1}(Z+|G|/\sigma > v/\sigma)\eta_q^{q-2}(|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(|G|/\sigma+Z;\chi))^2} = \mathbb{E}|G|^{q-2}. \tag{2.108}$$

Moreover, we can use similar arguments to obtain as $\sigma \to 0$

$$\begin{aligned}&\left|\int_{\mu}^{\infty}\int_{-\frac{g}{\sigma}+c_q\chi^{\frac{1}{2-q}}}^{-\frac{g}{\sigma}+\frac{v}{\sigma}}\frac{\sigma^{q-2}\eta_q^{q-2}(g/\sigma+z;\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(g/\sigma+z;\chi))^2}\phi(z)dzdF(g)\right|\\ \leq\ & C_0^{-2}\chi^{-1}(\eta_q^{+}(c_q;1))^{q-2}\sigma^{q-2}\int_{\mu}^{\infty}\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{-g/\sigma+v/\sigma}\phi(z)dzdF(g)\\ \leq\ & C_0^{-2}\chi^{-1}(\eta_q^{+}(c_q;1))^{q-2}\sigma^{q-2}(v/\sigma - c_q\chi^{\frac{1}{2-q}})\phi(-\mu/\sigma+v/\sigma) \to 0, \qquad (2.109)\end{aligned}$$

where the last inequality uses the fact that $g > \mu$ and

$$\int_{-g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{-g/\sigma+v/\sigma}\phi(z)dz < (v/\sigma - c_q\chi^{\frac{1}{2-q}})\phi(-g/\sigma+v/\sigma).$$

Putting together (2.107), (2.108) and (2.109) gives us

$$\lim_{\sigma\to 0}\frac{T_2}{\sigma^{2-q}} = -2\epsilon q(q-1)\mathbb{E}|G|^{q-2}.$$

Since T_3 and S_2 take similar forms as T_2, we can follow similar steps to derive,

$$\lim_{\sigma\to 0}\frac{T_3}{\sigma^{4-2q}\chi_q^*(\sigma)} = 2\epsilon q^2(q-1)(q-2)\mathbb{E}|G|^{2q-4}, \tag{2.110}$$

$$\lim_{\sigma\to 0}\frac{S_2}{\sigma^{2-2q}\chi_q^*(\sigma)} = 2\epsilon q^2\mathbb{E}|G|^{2q-2}. \tag{2.111}$$

Furthermore, by applying Lemma 2.5.36, it is not hard to see

$$\lim_{\sigma\to 0}\sigma^{q-2}T_1=0,\quad \lim_{\sigma\to 0}\sigma^{q-2}S_1=0. \tag{2.112}$$

Combing the results regarding T_1, T_2 and T_3 we have

$$\begin{aligned}\lim_{\sigma\to 0}\sigma^{q-2}S_3 &= \lim_{\sigma\to 0}\sigma^{q-2}T_1+\lim_{\sigma\to 0}\sigma^{q-2}T_2+\lim_{\sigma\to 0}\frac{T_3}{\sigma^{4-2q}\chi_q^*(\sigma)}\cdot\chi_q^*(\sigma)\sigma^{2-q}\\ &= -2\epsilon q(q-1)\mathbb{E}|G|^{q-2}. \end{aligned}\tag{2.113}$$

Collecting (2.110), (2.112) and (2.113) together, we obtain the order of $\frac{\partial R_2}{\partial\chi}$,

$$\lim_{\sigma\to 0}\frac{\partial R_2}{\partial\chi}/(\sigma^{2-2q}\chi_q^*(\sigma))=2\epsilon q^2\mathbb{E}|G|^{2q-2}. \tag{2.114}$$

From Equation (2.103), we observe that R_3 is only different from R_2 by a sign of g, hence we can follow the same arguments presented for analyzing $\partial R_2/\partial\chi$. We only highlight the differences for calculating T_2/σ^{2-q} (we are using the same notations):

1. $\lim_{\sigma\to 0}\frac{\mathbb{1}(Z-|G|/\sigma>v/\sigma)\eta_q^{q-2}(-|G|+\sigma Z;\sigma^{2-q}\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(-|G|/\sigma+Z;\chi))^2}=0,$
2. $\left|\int_\mu^\infty\int_{g/\sigma+c_q\chi^{\frac{1}{2-q}}}^{g/\sigma+v/\sigma}\frac{\sigma^{q-2}\eta_q^{q-2}(-g/\sigma+z;\chi)}{(1+\chi q(q-1)\eta_q^{q-2}(-g/\sigma+z;\chi))^2}\phi(z)dzdF(g)\right|$
   $\leq C_0^{-2}\chi^{-1}\eta_q^{q-2}(c_q;1)\sigma^{q-2}(v/\sigma-c_q\chi^{\frac{1}{2-q}})\phi(\mu/\sigma+c_q\chi^{\frac{1}{2-q}})=o(1)$.

Therefore, we can obtain $\lim_{\sigma\to 0}\frac{T_2}{\sigma^{2-q}}=0$ and conclude

$$\lim_{\sigma\to 0}\frac{\partial R_3}{\partial\chi}/(\sigma^{2-2q}\chi_q^*(\sigma))=0. \tag{2.115}$$

We combine the results from (2.103), (2.105), (2.106), (2.114), and (2.115) to have

$$\lim_{\sigma\to 0}\sigma^{2-2q}\chi_q^*(\sigma)2\epsilon q^2\mathbb{E}|G|^{2q-2}(2-q)\left[(\chi_q^*(\sigma))^{\frac{q+1}{2-q}}\phi(c_q(\chi_q^*(\sigma))^{\frac{1}{2-q}})2(1-\epsilon)c_q(\eta_q^+(c_q;1))^2\right]^{-1}=1.$$

Simplifying the above equality completes the proof. □

Lemma 2.5.38. *Suppose $\mathbb{P}(|G|>\mu)=1$ with μ being a fixed positive number and $\mathbb{E}|G|^2<\infty$, then for $0<q<1$ as $\sigma\to 0$,*

$$R_q(\chi_q^*(\sigma),\sigma)=\epsilon+\epsilon q^2\mathbb{E}|G|^{2q-2}(\chi_q^*(\sigma))^2\sigma^{2-2q}+o((\chi_q^*(\sigma))^2\sigma^{2-2q}).$$

Proof. We will use the same notation that was introduced in (2.103), and analyze the four terms respectively. Regarding R_2, we have

$$R_2 - \epsilon = \underbrace{\epsilon \int_{\mu}^{\infty} \int_{-g/\sigma + c_q \chi^{\frac{1}{2-q}}}^{\infty} (\eta_q(g/\sigma + z; \chi) - g/\sigma - z)^2 \phi(z) dz dF(g)}_{Q_1}$$
$$\underbrace{+ \epsilon \int_{\mu}^{\infty} \int_{-g/\sigma + c_q \chi^{\frac{1}{2-q}}}^{\infty} 2z(\eta_q(g/\sigma + z; \chi) - g/\sigma - z) \phi(z) dz dF(g)}_{Q_2}$$
$$\underbrace{-\epsilon \int_{\mu}^{\infty} \int_{-\infty}^{-g/\sigma + c_q \chi^{\frac{1}{2-q}}} z^2 \phi(z) dz dF(g)}_{Q_3}$$

According to Lemma 2.5.9 part (iii),

$$Q_1 = \epsilon q^2 \chi^2 \int_{\mu}^{\infty} \int_{-\frac{g}{\sigma} + c_q \chi^{\frac{1}{2-q}}}^{\infty} \eta_q^{2q-2}(\frac{g}{\sigma} + z; \chi) \phi(z) dz dF(g).$$

Using the same analysis of T_2 as in the proof of Lemma 2.5.37, it is clear that

$$\lim_{\sigma \to 0} \chi^{-2} \sigma^{2q-2} Q_1 = \epsilon q^2 \mathbb{E}|G|^{2q-2}.$$

Regarding Q_2, using integration by parts and Lemma 2.5.9 part (ii) we obtain

$$Q_2 = 2\epsilon(\eta_q^+(c_q \chi^{\frac{1}{2-q}}; \chi) - c_q \chi^{\frac{1}{2-q}}) \int_{\mu}^{\infty} \phi(-g/\sigma + c_q \chi^{\frac{1}{2-q}}) dF(g)$$
$$- \quad 2\epsilon \int_{\mu}^{\infty} \int_{-\frac{g}{\sigma} + c_q \chi^{\frac{1}{2-q}}}^{\infty} \frac{\chi q(q-1) \eta_q^{q-2}(u/\sigma + z; \chi)}{1 + \chi q(q-1) \eta_q^{q-2}(g/\sigma + z; \chi)} \phi(z) dz dF(g).$$

We can directly see the first term on the right hand side of the above equation is bounded by $O(\chi^{\frac{1}{2-q}} \phi(\mu/(2\sigma)))$. By using the same technique applied for analyzing T_2, we then know the second term is of order $\chi \sigma^{2-q}$. Hence we obtain

$$\lim_{\sigma \to 0} \chi^{-1} \sigma^{q-2} Q_2 = 2\epsilon q(1-q) \mathbb{E}|G|^{q-2}.$$

We now analyze Q_3. A simple integration by parts yields,

$$Q_3 = -\epsilon \int_{\mu}^{\infty} (g/\sigma - c_q \chi^{\frac{1}{2-q}}) \phi(g/\sigma - c_q \chi^{\frac{1}{2-q}}) dF(g) - \epsilon \int_{\mu}^{\infty} \int_{g/\sigma - c_q \chi^{\frac{1}{2-q}}}^{\infty} \phi(z) dz dF(g).$$

Using the fact that $\int_t^\infty \phi(z)dz \sim \frac{1}{t}\phi(t)$ and $\mu/\sigma - c_q\chi^{\frac{1}{2-q}} \to +\infty$, we can derive

$$\int_\mu^\infty (g/\sigma - c_q\chi^{\frac{1}{2-q}})\phi(g/\sigma - c_q\chi^{\frac{1}{2-q}})dF(g) \leq \int_\mu^\infty (g/\sigma)\phi(g/(2\sigma))dF(g) \leq \phi(\mu/(2\sigma))\mathbb{E}|G|,$$

and

$$\int_\mu^\infty \int_{g/\sigma - c_q\chi^{\frac{1}{2-q}}}^\infty \phi(z)dzdF(g) \leq \int_{\mu/\sigma - c_q\chi^{\frac{1}{2-q}}}^\infty \phi(z)dz \leq O(1/(\mu/\sigma - c_q\chi^{\frac{1}{2-q}})\phi(\mu/\sigma - c_q\chi^{\frac{1}{2-q}})).$$

It is then straightforward to confirm that $\lim_{\sigma\to 0}\chi^{-1}\sigma^{q-2}Q_3 = 0$. Combing the results about Q_1, Q_2 and Q_3 we obtain

$$\lim_{\sigma\to 0}\chi^{-2}\sigma^{2q-2}(R_2 - \epsilon) = \epsilon q^2\mathbb{E}|G|^{2q-2}. \tag{2.116}$$

Noticing that R_2 and R_3 take similar forms, the preceding arguments can be easily adapted to show

$$\lim_{\sigma\to 0}\chi^{-2}\sigma^{2q-2}R_3 = 0. \tag{2.117}$$

Regarding R_4, we first derive an upper bound in the following way:

$$\begin{aligned} R_4 = \epsilon\int_\mu^\infty \int_{-g/\sigma - c_q\chi^{\frac{1}{2-q}}}^{-g/\sigma + c_q\chi^{\frac{1}{2-q}}} \frac{g^2\phi(z)}{\sigma^2}dzdF(g) &\leq 2\epsilon c_q\chi^{\frac{1}{2-q}}\sigma^{-2}\int_\mu^\infty g^2\phi(-g/\sigma + c_q\chi^{\frac{1}{2-q}})dF(g) \\ &\leq 2\epsilon c_q\chi^{\frac{1}{2-q}}\sigma^{-2}\phi(-\mu/\sigma + c_q\chi^{\frac{1}{2-q}})\mathbb{E}|G|^2. \end{aligned}$$

The fact that $\sigma^{2-q}\chi \to 0$, as $\sigma \to 0$ leads to

$$\lim_{\sigma\to 0}\chi^{-2}\sigma^{2q-2}R_4 = 0. \tag{2.118}$$

We finally analyze R_1. A simple integration by parts gives us

$$\begin{aligned} &2(1-\epsilon)\int_{c_q\chi^{\frac{1}{2-q}}}^\infty \eta_q^2(z;\chi)\phi(z)dz = -2(1-\epsilon)\int_{c_q\chi^{\frac{1}{2-q}}}^\infty \frac{\eta_q^2(z;\chi)}{z}d\phi(z) = \\ &2(1-\epsilon)\frac{\eta_q^2(c_q\chi^{\frac{1}{2-q}};\chi)}{c_q\chi^{\frac{1}{2-q}}}\phi(c_q\chi^{\frac{1}{2-q}}) + 2(1-\epsilon)\int_{c_q\chi^{\frac{1}{2-q}}}^\infty \frac{2z\eta_q(z;\chi)\partial_1\eta_q(z;\chi) - \eta_q^2(z;\chi)}{z^2}\phi(z)dz. \end{aligned} \tag{2.119}$$

Since $|\eta_q(z;\chi)| \leq |z|$, we can bound the second integral in (2.119):

$$\left|\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty} \frac{2z\eta_q(z;\chi)\partial_1\eta_q(z;\chi)-\eta_q^2(z;\chi)}{z^2}\phi(z)dz\right| \leq \int_{c_q\chi^{\frac{1}{2-q}}}^{\infty} \frac{2}{|1+\chi q(q-1)\eta_q^{q-2}(z;\chi)|}\phi(z)dz$$
$$+\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}\phi(z)dz \overset{(1)}{\leq} \int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}\frac{2}{C}\phi(z)dz + \int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}\phi(z)dz \leq (2C^{-1}+1)\int_{c_q\chi^{\frac{1}{2-q}}}^{\infty}\phi(z)dz$$
$$\leq \quad O(\chi^{\frac{1}{q-2}}\phi(c_q\chi^{\frac{1}{2-q}})),$$

where (1) is due to Lemma 2.5.10. It is then clear that the dominant term in (2.119) is the first term and

$$\lim_{\sigma\to 0}\frac{R_1}{\chi^{\frac{1}{2-q}}\phi(c_q\chi^{\frac{1}{2-q}})} = \frac{2(1-\epsilon)(\eta_q^+(c_q;1))^2}{c_q}. \tag{2.120}$$

The results (2.116), (2.117), (2.118), (2.120) together with Lemma 2.5.37 finish the proof. □

We are in the position to derive the expansion in Theorem 2.2.10. According to Corollary 2.2.4,

$$\text{AMSE}(\lambda_{*,q},q,\sigma_w) = \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma}), \quad \bar{\sigma}^2 = \sigma_w^2 + \frac{\bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})}{\delta}. \tag{2.121}$$

First of all, implicit function theorem shows that $\bar{\sigma}$ is a continuous function of σ_w. Since $\bar{\sigma}=0$ when $\sigma_w=0$, we obtain $\bar{\sigma}\to 0$ as $\sigma_w\to 0$. Equation (2.121) combined with Lemma 2.5.38 yields

$$\lim_{\sigma_w\to 0}\frac{\bar{\sigma}^2}{\sigma_w^2} = \frac{\delta}{\delta-\epsilon}. \tag{2.122}$$

We now characterize the following limit:

$$\lim_{\sigma_w\to 0}\frac{\text{AMSE}(\lambda_{*,q},q,\sigma_w)-\frac{\delta\epsilon}{\delta-\epsilon}\sigma_w^2}{\sigma_w^{4-2q}(\chi_q^*(\bar{\sigma}))^2} \overset{(a)}{=} \lim_{\sigma_w\to 0}\frac{\bar{\sigma}^2\delta(R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})-\epsilon)}{\sigma_w^{4-2q}(\chi_q^*(\bar{\sigma}))^2(\delta-\epsilon)}$$
$$\overset{(b)}{=} \quad \frac{\delta}{\delta-\epsilon}\cdot\lim_{\bar{\sigma}\to 0}\frac{R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})-\epsilon}{(\chi_q^*(\bar{\sigma}))^2\bar{\sigma}^{2-2q}}\cdot\lim_{\bar{\sigma}\to 0}\frac{\bar{\sigma}^{4-2q}}{\sigma_w^{4-2q}} \overset{(c)}{=} \frac{\epsilon q^2\mathbb{E}|G|^{2q-2}\delta^{3-q}}{(\delta-\epsilon)^{3-q}}, \tag{2.123}$$

where (a) is due to (2.121), (b) holds since $\bar{\sigma}\to 0$ as $\sigma_w\to 0$, and (c) is obtained from (2.122) and Lemma 2.5.38. Our next step is to show $\chi_q^*(\bar{\sigma})\sim\chi_q^*(\sigma_w)$. Based on

Lemma 2.5.37 we obtain

$$\lim_{\sigma_w \to 0} \frac{\bar{\sigma}^{2-2q}}{(\chi_q^*(\bar{\sigma}))^{\frac{2q-1}{2-q}} \phi(c_q(\chi_q^*(\bar{\sigma}))^{\frac{1}{2-q}})} \cdot \frac{(\chi_q^*(\sigma_w))^{\frac{2q-1}{2-q}} \phi(c_q(\chi_q^*(\sigma_w))^{\frac{1}{2-q}})}{\sigma_w^{2-2q}} = 1.$$

It combined with (2.122) gives us

$$\lim_{\sigma_w \to 0} \frac{(\chi_q^*(\sigma_w))^{\frac{2q-1}{2-q}} \phi(c_q(\chi_q^*(\sigma_w))^{\frac{1}{2-q}})}{(\chi_q^*(\bar{\sigma}))^{\frac{2q-1}{2-q}} \phi(c_q(\chi_q^*(\bar{\sigma}))^{\frac{1}{2-q}})} = \left(1 - \frac{\epsilon}{\delta}\right)^{1-q},$$

which implies

$$\lim_{\sigma_w \to 0} \left[\frac{2q-1}{2-q} \log \frac{\chi_q^*(\sigma_w)}{\chi_q^*(\bar{\sigma})} - \frac{c_q^2}{2}(\chi_q^*(\sigma_w)^{2/(2-q)} - \chi_q^*(\bar{\sigma})^{2/(2-q)})\right] = (1-q)\log(1-\epsilon/\delta).$$

Since $\chi_q^*(\sigma_w), \chi_q^*(\bar{\sigma}) \to \infty$ as $\sigma_w \to 0$ from Lemma 2.5.36, dividing both sides of the above equality by $\chi_q^*(\sigma_w)^{2/(2-q)}$ yields

$$\lim_{\sigma_w \to 0} \frac{\chi_q^*(\sigma_w)}{\chi_q^*(\bar{\sigma})} = 1. \tag{2.124}$$

Finally, taking logarithm and diving by $(\chi_q^*(\sigma_w))^{\frac{2}{2-q}}$ on both sides of the equality in Lemma 2.5.37 leads to

$$\lim_{\sigma_w \to 0} \frac{(\chi_q^*(\sigma_w))^{\frac{2}{2-q}}}{\log \frac{1}{\sigma_w}} = \frac{4(1-q)}{c_q^2}. \tag{2.125}$$

Putting (2.123), (2.124) and (2.125) together completes the proof.

2.5.12 Proof of Theorem 2.2.11

The roadmap of the proof is similar to that of Theorem 2.2.10. We characterize the convergence rate of $\chi_q^*(\sigma)$ and derive the asymptotic formula for $R_q(\chi_q^*(\sigma), \sigma)$ in Lemma 2.5.39 and Lemma 2.5.40, respectively.

Lemma 2.5.39. *Suppose $\mathbb{E}|G|^2 < \infty$ and $\mathbb{P}(|G| > \mu) = 1$, where $\mu = \sup_v\{v : \mathbb{P}(|G| > v) = 1\} > 0$. Then for $q = 0$,*

$$\lim_{\sigma \to 0} \sqrt{\chi_q^*(\sigma)}\sigma = \frac{\mu}{2c_q}.$$

Proof. By using the same arguments presented in the proof of Lemma 2.5.36, we can obtain $\chi_q^*(\sigma) \to \infty$, as $\sigma \to 0$. Now we consider an arbitrarily convergent sequence $\sigma_k \to 0$, as $k \to \infty$, and show $\sqrt{\chi_q^*(\sigma_k)}\sigma_k \to \mu/(2c_0)$. Denote $\lim_{k\to\infty} \sqrt{\chi_q^*(\sigma_k)}\sigma_k = \alpha$. For notational simplicity, below we may write χ for $\chi_q^*(\sigma)$. Suppose $\alpha > \mu/c_0$, then by Fatou's lemma, we have

$$\liminf_{k\to\infty} \mathbb{E}(\eta_0(G/\sigma_k + Z;\chi) - G/\sigma_k)^2 \geq \liminf_{k\to\infty} \mathbb{E}[\mathbb{1}(|G+\sigma_k Z| \leq c_0\sqrt{\chi}\sigma_k)G^2/\sigma_k^2] = \infty.$$

On the other hand, $R_0(\chi, \sigma_k) \leq R_0(0, \sigma_k) = 1$. This is a contradiction. Hence it holds that $\alpha \leq \mu/c_0$. Next we aim to show $\alpha \leq \mu/(2c_0)$. Since $\eta_0(u;\chi) = u\mathbb{1}(|u| > c_0\sqrt{\chi})$, it is straightforward to confirm the following

$$\begin{aligned} R_0(\chi,\sigma) = {} & \underbrace{2(1-\epsilon)\Big[c_0\sqrt{\chi}\phi(c_0\sqrt{\chi}) + \int_{c_0\sqrt{\chi}}^{\infty}\phi(z)dz\Big]}_{R_1} \\ + {} & \underbrace{\epsilon\mathbb{E}\Bigg[\Big(c_0\sqrt{\chi} - \frac{|G|}{\sigma}\Big)\phi\Big(c_0\sqrt{\chi} - \frac{|G|}{\sigma}\Big)}_{R_2} + \int_{c_0\sqrt{\chi}-\frac{|G|}{\sigma}}^{\infty}\phi(z)dz\Bigg] \\ + {} & \underbrace{\epsilon\mathbb{E}\Bigg[\Big(c_0\sqrt{\chi} + \frac{|G|}{\sigma}\Big)\phi\Big(c_0\sqrt{\chi} + \frac{|G|}{\sigma}\Big) + \int_{c_0\sqrt{\chi}+\frac{|G|}{\sigma}}^{\infty}\phi(z)dz\Bigg]}_{R_3} \\ + {} & \underbrace{\epsilon\mathbb{E}\int_{-c_0\sqrt{\chi}-\frac{|G|}{\sigma}}^{c_0\sqrt{\chi}-\frac{|G|}{\sigma}} \frac{|G|^2}{\sigma^2}\phi(z)dz}_{R_4}. \end{aligned} \tag{2.126}$$

Moreover, it is clear that $\chi_q^*(\sigma_k)$, the optimal thresholding value, is finite and non-zero, and hence we have $\frac{\partial R_0(\chi_q^*(\sigma_k),\sigma_k)}{\partial \chi} = 0$, i.e.,

$$\frac{\partial R_1}{\partial \chi} + \frac{\partial R_2}{\partial \chi} + \frac{\partial R_3}{\partial \chi} + \frac{\partial R_4}{\partial \chi} = 0, \tag{2.127}$$

where we can calculate the partial derivatives as follows[6]

$$\begin{aligned}
\frac{\partial R_1}{\partial \chi} &= (\epsilon-1)c_0^3\sqrt{\chi}\phi(c_0\sqrt{\chi}),\\
\frac{\partial R_2}{\partial \chi} &= \frac{-\epsilon c_0}{2\sqrt{\chi}}\mathbb{E}\big[\big(c_0\sqrt{\chi}-|G|/\sigma_k\big)^2\phi\big(c_0\sqrt{\chi}-|G|/\sigma_k\big)\big],\\
\frac{\partial R_3}{\partial \chi} &= \frac{-\epsilon c_0}{2\sqrt{\chi}}\mathbb{E}\big[\big(c_0\sqrt{\chi}+|G|/\sigma_k\big)^2\phi\big(c_0\sqrt{\chi}+|G|/\sigma_k\big)\big],\\
\frac{\partial R_4}{\partial \chi} &= \frac{\epsilon c_0}{2\sqrt{\chi}\sigma_k^2}\mathbb{E}\left\{|G|^2\left[\phi\big(c_0\sqrt{\chi}-|G|/\sigma_k\big)+\phi\big(c_0\sqrt{\chi}+|G|/\sigma_k\big)\right]\right\}.
\end{aligned}$$

With a few more steps of calculations we obtain as $\sigma_k \to 0$,

$$\begin{aligned}
&\frac{\partial R_2}{\partial \chi}+\frac{\partial R_3}{\partial \chi}+\frac{\partial R_4}{\partial \chi} = \frac{-\epsilon c_0^2}{2\sigma_k}\mathbb{E}\big[(c_0\sqrt{\chi}\sigma_k-2|G|)\phi\big(c_0\sqrt{\chi}-|G|/\sigma_k\big)\big]\\
&+\frac{-\epsilon c_0^2}{2\sigma_k}\mathbb{E}\big[\big(c_0\sqrt{\chi}\sigma_k+2|G|)\phi\big(c_0\sqrt{\chi}+|G|/\sigma_k\big)\big] \propto \frac{1}{\sigma_k}\mathbb{E}\big[|G|\phi\big(c_0\sqrt{\chi}-|G|/\sigma_k\big)\big].
\end{aligned}$$

Hence, dividing both sides of (2.127) by $\sqrt{\chi}\phi(c_0\sqrt{\chi})$ and letting $k\to\infty$ shows

$$0 < \lim_{k\to\infty}\mathbb{E}\big[|G|\exp\big(|G|(-|G|+2\sigma_k c_0\sqrt{\chi})/(2\sigma_k^2)\big)\big] < \infty. \tag{2.128}$$

If $\alpha > \mu/(2c_0)$, we will see

$$\begin{aligned}
&\mathbb{E}\big[|G|\exp\big(|G|(-|G|+2\sigma_k c_0\sqrt{\chi})/(2\sigma_k^2)\big)\big]\\
\geq\ &\mathbb{E}\big[|G|\exp\big(|G|(-|G|+2\sigma_k c_0\sqrt{\chi})/(2\sigma_k^2)\big)\cdot\mathbb{1}(|G|<2\alpha c_0)\big] \to +\infty.
\end{aligned}$$

We have used Fatou's lemma to obtain the last limit. Obviously the inequality above contradicts (2.128). Thus we obtain the upper bound $\mu/(2c_0)$ for α. Finally we would like to derive $\alpha \geq \mu/(2c_0)$. Note that since $\alpha \leq \mu/(2c_0)$, it is not hard to show that when k is large,

$$\frac{\partial R_4}{\partial \chi} \leq \frac{\epsilon c_0\mathbb{E}|G|^2}{\sqrt{\chi}\sigma_k^2}\phi(c_0\sqrt{\chi}-\mu/\sigma_k) = O\big((\sqrt{\chi}\sigma_k^2)^{-1}\phi(c_0\sqrt{\chi}-\mu/\sigma_k)\big).$$

Based on the inequality above, we can further obtain

$$\left|\frac{\partial R_2}{\partial \chi}+\frac{\partial R_3}{\partial \chi}+\frac{\partial R_4}{\partial \chi}\right| \leq O\big((\sqrt{\chi}\sigma_k^2)^{-1}\phi(c_0\sqrt{\chi}-\mu/\sigma_k)\big). \tag{2.129}$$

[6]The condition $\mathbb{E}|G|^2<\infty$ enables us to apply dominated convergence theorem to exchange the differentiation and expectation in the calculation of the partial derivatives.

Now suppose $\alpha < \mu/(2c_0)$, then it follows that

$$\frac{1}{\sqrt{\chi}\sigma_k^2}\phi\Big(c_0\sqrt{\chi}-\frac{\mu}{\sigma_k}\Big)\cdot\frac{1}{\sqrt{\chi}\phi(c_0\sqrt{\chi})}=\frac{1}{\chi\sigma_k^2}\exp\Big(\frac{\mu(\mu-2c_0\sigma_k\sqrt{\chi})}{-2\sigma_k^2}\Big)=o(1).$$

However, this fact combined with (2.129) implies that if we divide both sides of Equation (2.127) by $\sqrt{\chi}\phi(c_0\sqrt{\chi})$ and letting $k\to\infty$, we would get

$$(\epsilon-1)c_0^3=0,$$

which is a contradiction. Above all we have proved that for an arbitrarily convergent sequence $\sigma_k\to 0$, $\sqrt{\chi_q^*(\sigma_k)}\sigma_k\to\mu/(2c_0)$, as $k\to\infty$. This completes the proof. □

Lemma 2.5.40. *Suppose* $\mathbb{E}|G|^2<\infty$ *and* $\mathbb{P}(|G|>\mu)=1$*, where* $\mu=\sup_v\{v:\mathbb{P}(|G|>v)=1\}>0$*. Then for* $q=0$*, as* $\sigma\to 0$

$$R_q(\chi_q^*(\sigma),\sigma)=\epsilon+O\big(\sqrt{\chi_q^*(\sigma)}\phi(c_0\sqrt{\chi_q^*(\sigma)})\big).$$

Proof. We adopt the same notations from the proof of Lemma 2.5.39. Then,

$$R_q(\chi_q^*(\sigma),\sigma)-\epsilon=R_1+(R_2-\epsilon)+R_3+R_4.$$

Based on the result that $\sqrt{\chi_q^*(\sigma)}\sigma\to\frac{\mu}{2c_0}$ in Lemma 2.5.39, from (2.126) we can obtain

$$\begin{aligned}
R_1 &= O\big(\sqrt{\chi_q^*(\sigma)}\phi(c_0\sqrt{\chi_q^*(\sigma)})\big),\\
R_2-\epsilon &= O\big(\mathbb{E}\big[|G|/\sigma\phi\big(c_0\sqrt{\chi_q^*(\sigma)}-|G|/\sigma\big)\big]\big),\\
R_3 &= O\big(\mathbb{E}\big[|G|/\sigma\phi\big(c_0\sqrt{\chi_q^*(\sigma)}+|G|/\sigma\big)\big]\big).
\end{aligned}$$

Regarding R_4, it holds that

$$\begin{aligned}
R_4 &= \epsilon\mathbb{E}\Big[|G|^2/\sigma^2\cdot\Big(\int_{|G|/\sigma-c_0\sqrt{\chi_q^*(\sigma)}}^{\infty}\phi(z)dz-\int_{|G|/\sigma+c_0\sqrt{\chi_q^*(\sigma)}}^{\infty}\phi(z)dz\Big)\Big]\\
&\le \epsilon\mathbb{E}\Big[|G|^2/\sigma^2\cdot\int_{|G|/\sigma-c_0\sqrt{\chi_q^*(\sigma)}}^{\infty}\phi(z)dz\Big]\\
&\le \epsilon\mathbb{E}\Big[|G|/\sigma\cdot\phi(|G|/\sigma-c_0\sqrt{\chi_q^*(\sigma)})\cdot|G|/(|G|-c_0\sigma\sqrt{\chi_q^*(\sigma)})\Big]\\
&\le O\big(\mathbb{E}\big[|G|/\sigma\cdot\phi(|G|/\sigma-c_0\sqrt{\chi_q^*(\sigma)})\big]\big).
\end{aligned}$$

Furthermore, (2.128) shows us that

$$\mathbb{E}\Big[|G|/\sigma\phi(|G|/\sigma - c_0\sqrt{\chi_q^*(\sigma)})\Big] \cdot \frac{1}{\sqrt{\chi_q^*(\sigma)}\phi(c_0\sqrt{\chi_q^*(\sigma)})} = O(1).$$

Putting together what we have derived so far closes the proof. $\square$

Based on Lemmas 2.5.39 and 2.5.40, it is straightforward to obtain the following result:

$$R_q(\chi_q^*(\sigma), \sigma) = \epsilon + o(\phi(\tilde{\mu}\sigma^{-1})).$$

The expansion of $\mathrm{AMSE}(\lambda_{*,0}, 0, \sigma_w)$ can be derived accordingly as we did for $0 < q < 1$. We do not repeat the arguments.

2.5.13 Proof of Theorem 2.2.12

The proof of Theorem 2.2.12 is similar to that of Theorem 2.2.6. We hence skip it for brevity.

Chapter 3

Low noise analysis without sparsity

In Chapter 2, we have discussed the limitations of phase transition diagrams and performed a second-order low noise sensitivity analysis to resolve such issue. The fundamental condition underlying all the preceding analyses is the sparsity of the coefficient β. However, exact sparsity might be a stringent requirement from practical point of view. A more realistic assumption is that β is approximately sparse, i.e., it has many elements of small values. Then how would the bridge regression estimators behave? As we shall see in this chapter, phase transition analysis is not sufficient to characterize the performance of LQLS estimators, and instead we present the low noise sensitivity analysis, as a generalization of the phase transition, to provide a more accurate view of LQLS for estimating non-sparse β. To simplify the presentation, we focus our analysis on the case $1 \leq q \leq 2$.

3.1 Introduction

3.1.1 Objective

Phase transition analysis (PT) studies the asymptotic mean square error (AMSE) $\|\hat{\beta}(\lambda, q) - \beta\|_2^2/p$ under the asymptotic setting $p \to \infty$ and $n/p \to \delta$. Then, it considers $w = 0$ and calculates the smallest δ for which $\inf_\lambda \lim_{p\to\infty} \|\hat{\beta}(\lambda, q) - \beta\|_2^2/p = 0$. In this chapter, we consider situations in which β is not exactly sparse. As is intuitively expected and will be discussed later in the chapter, the phase transition analysis implies that for every $1 \leq q \leq 2$, if $\delta > 1$, then $\inf_\lambda \lim_{p\to\infty} \|\hat{\beta}(\lambda, q) - \beta\|_2^2/p = 0$ and if $\delta < 1$, then $\inf_\lambda \lim_{p\to\infty} \|\hat{\beta}(\lambda, q) - \beta\|_2^2/p \neq 0$. This simple application of PT reveals some of the limitations of the phase transition analysis:

1. Phase transition analysis is concerned with $w = 0$, and when β is not sparse, LQLS with different values of q have the same phase transition at $\delta = 1$. The same phase transition happens for ordinary least squares (OLS). Hence, it is not clear whether regularization can improve the performance of OLS and if it does, which regularizer is the best. We expect the choice of regularizer to matter when we add some noise to the measurements.

2. Phase transition diagram is not sensitive to the magnitude distribution of the elements of β. Again, intuitively speaking, this seems to have a major impact on the performance of different estimators when the noise is present in the system.

Following the same idea presented in Chapter 2, we perform a second-order low noise sensitivity analysis to overcome the limitations. This framework has the following two main advantages over the phase transition analysis:

1. It reveals certain phenomena that are important in applications and are not captured by PT analysis. For instance, one immediately sees the impact of the regularizer and the magnitude distribution of the elements of β on the AMSE. Furthermore, these relations are expressed explicitly and can be interpreted easily.

2. It provides a bridge between the phase transition analysis proposed in compressed sensing, and the classical large sample-size asymptotics ($n/p \to \infty$). We will discuss some of the implications of this connection for the classical asymptotics in Chapter 3.3.

To demonstrate the above claims we use the low-noise sensitivity analysis to address the following questions:

1. When β is not sparse, does LASSO outperform LQLS with $q \in (1, 2]$? Which LQLS performs the best?
2. What is the impact of the distribution of the elements of β on AMSE of LQLS estimators?

3.1.2 Related works

The works most related to ours are Donoho (2006a); Candes and Tao (2006); Candes and Plan (2011). In the first two papers, the authors considered non-sparse β with the constraint that $\|\beta\|_q \leq R$ or the ith largest component $|\beta|_{(i)}$ decays as $i^{-\alpha}(\alpha > 0)$. The papers derived optimal (up to logarithmic factor) upper bounds on the mean square error of LASSO. However, we characterize the performance of LASSO for a generic β and derive conditions under which LASSO outperforms other bridge estimators. Also, we should emphasize that thanks to our asymptotic settings, unlike these two papers we are able to derive exact expressions of AMSE with sharp constants. Finally, Candes and Plan (2011) studied a fixed signal β and obtained an oracle inequality for $\|\hat{\beta}(\lambda, 1) - \beta\|_2$, with the tuning λ chosen as an explicit function of p. While their results are more general than ours, the bounds suffer from loose constants and are not sufficient to provide sharp comparison of LASSO with other LQLS. Moreover, the tuning parameter λ in our case is set to the optimal one that minimizes the AMSE for every LQLS, which further paves our way for accurate comparison between different LQLS.

The performance of LQLS with $q \geq 0$ under classical asymptotic setting where p is fixed and $n \to \infty$ is studied in Knight and Fu (2000). The author obtained the $\sqrt{n}$ convergence of LQLS estimates and derived the asymptotic distributions. His results can be used to calculate the AMSE for LQLS with optimal tuning and show that they are all equal for $q \in [1, 2]$. However, we demonstrate in Chapter 3.3 that by a second-order analysis, a more accurate comparison between the performances of different LQLS is possible. In particular, LASSO will be shown to outperform others for certain type of non-sparse coefficients.

3.2 Phase transition and a second-order noise sensitivity analysis

Recall the asymptotic framework we introduced in Chapter 1.2. In the rest of this chapter, we will assume f_β does not have any point mass at zero. We use the notation B to denote a one dimensional random variable distributed according to f_β.

3.2.1 Phase transition

Suppose that there is no noise in the linear model, i.e., $\sigma_w = 0$. Our first goal in the phase transition analysis is to find the minimum value of δ for which $\text{AMSE}(\lambda_{*,q}, q, 0) = 0$. Our next theorem characterizes the phase transition.

Theorem 3.2.1. *Let $q \in [1, 2]$. If $\mathbb{E}|B|^2 < \infty$, then we have*

$$\text{AMSE}(\lambda_{*,q}, q, 0) = \begin{cases} > 0 & \text{if } \delta < 1, \\ = 0 & \text{if } \delta > 1. \end{cases}$$

The result can also be derived from several different frameworks including the statistical dimension framework in Amelunxen et al. (2014). But we derive it as a simple byproduct of our results in Chapter 3.2.2. So we do not discuss its proof here. This result is not surprising. Since, none of the coefficients is zero, the exact recovery is impossible if $n < p$. Also, note that when $\delta > 1$ even the ordinary least squares is capable of recovering β. Hence, the result of phase transition analysis does not provide any additional information on the performance of different regularizers. It is not even capable of showing the advantage of regularization techniques over the standard least squares algorithm. This is due to the fact that the result of Theorem 3.2.1 holds only in the noiseless case. Intuitively speaking, in the practical settings where the existence of the measurement noise is inevitable, we expect different LQLS to behave differently. For instance, even though the coefficient under study is not sparse, when f_β has a large mass around zero (it is approximately sparse), we expect the sparsity promoting LASSO to offer better performance than the other LQLS. However, the distribution f_β does not have any effect on the phase transition diagram. Motivated by these concerns, in the next section, we investigate the performance of LQLS in the noisy setting, and study their noise sensitivity when the noise level σ_w is small. The new analysis will offer more informative answers.

3.2.2 Second-order noise sensitivity analysis of AMSE

As an immediate generalization of the phase transition analysis, we can study the performance of different estimators in the presence of a small amount of noise. More formally, we derive the asymptotic expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ for every $q \in [1, 2]$, when $\sigma_w \to 0$. As will be discussed later, this generalization of phase transitions presents a more delicate analysis of LQLS. We start with the study of AMSE for the ordinary least squares (OLS). The result of OLS will be later used for comparison purposes.

Corollary 3.2.2. *Consider the region $\delta > 1$. For the OLS estimate $\hat{\beta}(0, q)$, we have*

$$\text{AMSE}(0, q, \sigma_w) = \frac{\sigma_w^2}{1 - 1/\delta}.$$

We prove the above corollary in Chapter 3.4.2. Note that the proof we presented there, has not used the independence of the noise elements that is often assumed in the analysis of OLS. Now we can discuss LQLS with the optimal choice of λ. We first consider the coefficients whose elements are bounded away from zero in Theorem 3.2.3 and then study other distributions in Theorem 3.2.5.

Theorem 3.2.3. *Consider the region $\delta > 1$. Suppose that $\mathbb{P}(|B| > \mu) = 1$ with μ being a positive constant and $\mathbb{E}|B|^2 < \infty$. Then, for $q \in (1, 2]$, as $\sigma_w \to 0$*

$$\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1 - 1/\delta} - \frac{\delta^3 (q-1)^2 (\mathbb{E}|B|^{q-2})^2}{(\delta - 1)^3 \mathbb{E}|B|^{2q-2}} \sigma_w^4 + o(\sigma_w^4),$$

and for $q = 1$, as $\sigma_w \to 0$

$$\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1 - 1/\delta} - |o(e^{-\frac{\tilde{\mu}^2(\delta-1)}{\delta \sigma_w^2}})|,$$

where $\tilde{\mu}$ is any positive number smaller than μ.

The proof can be found in Chapter 3.4.3. We observe that the first dominant term in the expansion of $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$ is exactly the same for all values of q, including $q = 1$ and is equal to $\sigma_w^2/(1 - 1/\delta)$. This is also the same as the AMSE of the OLS. We may consider this term as the "phase transition" term, since it will go to zero only when $\delta > 1$. In a nutshell, the first term in the expansion provides the phase transition information. However, we are able to derive the second order term for $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$. This term gives us what is beyond phase transition analysis. The impact of the signal distribution f_β and the regularizer ℓ_q, that is omitted in PT diagram, is revealed in the second order term. As a result, to compare the performance of LQLS with different values of q in the low noise regime, we can compare their second order terms.

First note that all the regularizers that are studied in Theorem 3.2.3 improve the performance of OLS. When the distribution of the coefficients is bounded away from 0, no significant gain is obtained from LASSO since the second dominant term in the expansion of AMSE is exponentially small. However, the rate of the second order term exhibits an interesting transition from exponential to a polynomial decay when q increases from 1. In fact, it seems that bridge regularizers with $q > 1$ offer more substantial improvements over OLS. Even though LASSO is suboptimal, it is not clear which value of q provides the best performance here. Among other LQLS with $q \in (1, 2]$, the optimality is determined by the constant involved in the second order term (they all have the same orders). To simplify our discussions, define

$$C_q = \frac{(q-1)^2(\mathbb{E}|B|^{q-2})^2}{\mathbb{E}|B|^{2q-2}}, \quad q^* = \arg\max_{1<q\leq 2} C_q.$$

Then LQLS with $q = q^*$ will perform the best. To provide some insights on q^*, we focus on a special family of distributions.

Lemma 3.2.4. *Consider the two-point mixture* $|B| \sim \alpha\Delta_{\mu_1} + (1-\alpha)\Delta_{\mu_2}$ *with* $0 < \mu_1 \leq \mu_2, \alpha \in (0,1)$. *Then* $q^* = 2$ *when* $\mu_1 = \mu_2$, *and* $q^* \to 1$ *as* $\mu_2/\mu_1 \to \infty$.

Proof. When $\mu_1 = \mu_2$, it is clear that $C_q = (q-1)^2\mu_1^{-2}$ and thus $q^* = 2$. We now consider $0 < \mu_1 < \mu_2$. Denote $\kappa = \mu_2/\mu_1$. We can then write C_q as:

$$C_q = \frac{(q-1)^2(\alpha\mu_1^{q-2} + (1-\alpha)\mu_2^{q-2})^2}{\alpha\mu_1^{2q-2} + (1-\alpha)\mu_2^{2q-2}} = \frac{(q-1)^2(\alpha + (1-\alpha)\kappa^{q-2})^2}{\mu_1^2(\alpha + (1-\alpha)\kappa^{2q-2})}.$$

Define $\bar{q} = 1 + \frac{1}{\log\kappa}$. We would like to show that for any $\epsilon > 0$, $C_{\bar{q}} > \max_{1+\epsilon\leq q\leq 2} C_q$ for κ large enough. That will give us $q^* \in (1, 1+\epsilon)$ and hence finishes the proof. To show that, note that for any $q \in [1+\epsilon, 2], \kappa \geq 1$,

$$\begin{aligned} \frac{C_{\bar{q}}}{C_q} &= \frac{(\alpha + (1-\alpha)\kappa^{\bar{q}-2})^2}{(\alpha + (1-\alpha)\kappa^{q-2})^2} \cdot \frac{\alpha + (1-\alpha)\kappa^{2q-2}}{\alpha + (1-\alpha)\kappa^{2\bar{q}-2}} \cdot \frac{(\bar{q}-1)^2}{(q-1)^2} \\ &\geq \alpha^2 \cdot \frac{\alpha + (1-\alpha)\kappa^{2\epsilon}}{\alpha + (1-\alpha)e^2} \cdot (\bar{q}-1)^2 \geq \frac{\alpha^2(1-\alpha)}{\alpha + (1-\alpha)e^2} \cdot \kappa^{2\epsilon}(\log\kappa)^{-2} \to \infty. \end{aligned}$$

Therefore, $C_{\bar{q}} > \max_{1+\epsilon\leq q\leq 2} C_q$ when κ is sufficiently large. □

Lemma 3.2.4 implies that ridge ($q = 2$) regularizer is optimal when the two-point mixture components coincide, and the optimal value of q will shift towards 1 as the ratio of the two points goes off to infinity. Intuitively speaking, one would expect ridge to penalize large coefficients more aggressively than $q < 2$. Hence, in cases the coefficient has a large dynamic range, ridge penalizes the large coefficient values more and is not expected to outperform other values of q. Note that for the two-point mixture coefficients, the optimal value of q can be arbitrarily close to 1, however LASSO can never be optimal because its second order term is exponentially small. We next study a more informative and interesting case where the distribution of β has more mass around zero.

Theorem 3.2.5. *Consider the region $\delta > 1$ and assume $\mathbb{E}|B|^2 < \infty$. For any given $q \in (1,2)$, suppose that $\mathbb{P}(|B| \leq t) = O(t^{2-q+\epsilon})$ (as $t \to 0$) with ϵ being any positive constant, then as $\sigma_w \to 0$*

$$\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) = \frac{\sigma_w^2}{1-1/\delta} - \frac{\delta^3 (q-1)^2 (\mathbb{E}|B|^{q-2})^2}{(\delta-1)^3 \mathbb{E}|B|^{2q-2}} \sigma_w^4 + o(\sigma_w^4).$$

For $q = 2$, as $\sigma_w \to 0$

$$\mathrm{AMSE}(\lambda_{*,2}, 2, \sigma_w) = \frac{\sigma_w^2}{1-1/\delta} - \frac{\delta^3}{(\delta-1)^3 \mathbb{E}|B|^2} \sigma_w^4 + o(\sigma_w^4).$$

For $q = 1$, suppose that $\mathbb{P}(|B| \leq t) = \Theta(t^\ell)$ with $\ell > 0$, then as $\sigma_w \to 0$

$$-|\Theta(\sigma_w^{2\ell+2})| \cdot \Bigg(\underbrace{\log \log \ldots \log}_{m \text{ times}} \left(\frac{1}{\sigma_w} \right) \Bigg)^{\ell} \lesssim \mathrm{AMSE}(\lambda_{*,1}, 1, \sigma_w) - \frac{\sigma_w^2}{1-1/\delta} \lesssim -|\Theta(\sigma_w^{2\ell+2})|,$$

where m can be any natural number.

The proof is presented in Chapter 3.4.4. Note that the condition $\mathbb{P}(|B| \leq t) = O(t^{2-q+\epsilon})$ for $q \in (1,2)$ is necessary otherwise the form $\mathbb{E}|B|^{q-2}$ appearing in the second order term will be unbounded. We would like to make the following observations:

1. Compared to the results in Theorem 3.2.3, we see that the expansion of AMSE for $q \in (1,2]$ in Theorem 3.2.5 remains the same for more general B, while the rate of the second order term for LASSO changes to polynomial from exponential. That means LASSO is more sensitive to the distribution of β than other LQLS.

2. The second order term of LASSO becomes smaller as ℓ decreases. It implies that LASSO performs better when the probability mass of the coefficient concentrates more around zero. This can be well explained by the sparsity promoting feature of LASSO;

3. As in the case $\mathbb{P}(|B| > \mu) = 1$, the first dominant term is the same for all $q \in [1,2]$. Hence we have to compare their second order term. For any

given $q \in (1, 2]$, suppose $\mathbb{P}(|B| \leq t) = \Theta(t^{2-q+\epsilon})$, then the second term of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ is of order σ_w^4, while that of LASSO is $\Theta(\sigma_w^{6-2q+2\epsilon})$ (ignore the logarithmic factor). Since both terms are negative, we can conclude LASSO performs better than LQLS with that value of q when $\epsilon \in (0, q-1)$, and performs worse when $\epsilon \in (q-1, \infty)$. This observation has an important implication. The behavior of the distribution of $|B|$ around zero is the most important factor in the comparison between LASSO and other LQLS. If the probability density function (pdf) of B is zero at zero, then we should not use LASSO and when it goes to infinity LASSO performs better than LQLS with $q > 1$ (at least for those values of q for which our theorem is applicable).

4. Regrading the case where the pdf is finite at zero, our calculations of LASSO are not sharp enough to give an accurate comparison between LASSO and other LQLS. However, the comparison of LQLS for different values of $q > 1$ will shed more light on the performance of different regularizers in this case. Hence, we consider one of the most popular families of distributions and present an accurate comparison among $q \in (1, 2]$.

Lemma 3.2.6. *Consider $|B|$ with density function $f(b) = \zeta(\tau, q_0)e^{-\tau b^{q_0}}\mathbb{1}(0 \leq b < \infty)$, where $q_0 \in (0, 2], \tau > 0$ and $\zeta(\tau, q_0)$ is the normalization constant. Then $q^* = \max(1, q_0)$.*

Proof. A simple integration by parts yields, for $q \in (1, 2]$

$$\mathbb{E}|B|^{q-2} = \int_0^\infty \zeta(\tau, q_0) b^{q-2} e^{-\tau b^{q_0}} db = \frac{\tau q_0}{q-1} \int_0^\infty \zeta(\tau, q_0) b^{q+q_0-2} e^{-\tau b^{q_0}} db = \frac{\tau q_0 \mathbb{E}|B|^{q+q_0-2}}{q-1}.$$

Hence $C_q = \tau^2 q_0^2 \frac{(\mathbb{E}|B|^{q+q_0-2})^2}{\mathbb{E}|B|^{2q-2}}$. We first consider $q_0 \in [1, 2]$, then

$$C_q = \tau^2 q_0 \frac{[\mathbb{E}(|B|^{q-1} \cdot |B|^{q_0-1})]^2}{\mathbb{E}|B|^{2q-2}} \overset{(a)}{\leq} \tau^2 q_0 \mathbb{E}|B|^{2q_0-2} = C_{q_0},$$

where (a) is due to Cauchy-Schwarz inequality. So we obtain $q^* = q_0$. Regrading the case $q_0 \in (0, 1)$, let B' be an independent copy of B. Then for any $q \in [1, 2]$

$$\mathbb{E}|B|^{q+q_0-2} - \mathbb{E}|B|^{q-1}\mathbb{E}|B|^{q_0-1} = \frac{1}{2}\mathbb{E}(|B|^{q-1} - |B'|^{q-1})(|B|^{q_0-1} - |B'|^{q_0-1}) \leq 0.$$

As a result, we can derive

$$C_q \leq \tau^2 q_0 \frac{(\mathbb{E}|B|^{q+q_0-2})^2}{(\mathbb{E}|B|^{q-1})^2} \leq \tau^2 q_0 \mathbb{E}(|B|^{q_0-1})^2 = C_1.$$

We can then conclude $q^* = 1 = \max(1, q_0)$. □

As we discussed after Theorem 3.2.5, the shape of the coefficient distribution around zero is the most important factor when it comes to the comparison of different ℓ_q-regularizers. Lemma 3.2.6 indicates that among the distributions whose pdf exists and is non-zero at zero, the tail behavior has an influence on the performance of LQLS. In particular, LQLS with $q = q_0 \in (1, 2]$ is optimal for distributions with the exponential decay tail $e^{-\tau b^{q_0}}$. Since $\hat{\beta}(\lambda, q_0)$ can be considered as the maximum a posterior estimate (MAP), our result suggests that MAP offers the best performance in the low noise regime (among the bridge estimators). This is in general not true. See Zheng et al. (2017) for a counterexample in large noise cases. It is also interesting to observe that as the tail becomes heavier than that of Laplacian distribution, the optimal q^* approaches 1. Again note, that this observation is consistent with the fact that ridge often penalizes large coefficient values more aggressively than the other estimators. Hence, if the tail of the distribution is light (like Gaussian distributions), then ridge offers the best performance, otherwise, other values of q offer better results.

Based on Theorems 3.2.3, 3.2.5 and follow-up discussions, we are ready to summarize the answers to the two questions we target in Chapter 3.1.1. In the high signal-to-noise ratio regime, we can conclude that

1. How LASSO compares with other LQLS largely depends on the distribution of the coefficient. The behavior of the distribution around zero is the most important factor. When the probability density of the coefficient at zero is finite, then the tail behavior of the distribution plays a role too.

2. LQLS with $q \in (1, 2]$ outperforms LASSO when the distribution of the coefficient is bounded away from zero. For two-point mixture distributions, ridge is

optimal if the two points overlap and the optimal q^* approaches 1 as the two points move away from each other. When B has probability mass around zero, $1 < q \leq 2$ can still beat LASSO if $\mathbb{P}(|B| \leq t) = \Theta(t^{2-q+\epsilon})$ with $\epsilon > q - 1$. LASSO starts to outperform other values of q when $\epsilon \in (0, q-1)$. For the distribution with tail $e^{-\tau b^{q_0}}(1 < q_0 \leq 2)$, LQLS with $q = q_0$ is optimal among $q \in (1, 2]$.

3.3 Implications for classical asymptotics

Our analysis so far has been focused on the high-dimensional setting in which $n/p \to \delta \in (0, \infty)$. Furthermore, we assumed that the noise variance is small. At an intuitive level this platform seems to be connected to the classical asymptotic framework that has been studied in statistics extensively. In the classical asymptotics, it is assumed that the signal-to-noise ratio of each observation is fixed and $n/p \to \infty$. Note that having more measurements is at the intuitive level equivalent to less noise. Hence, we expect our low-noise sensitivity to have some implications for the classical asymptotics too. Our goal below is to formalize this connection and explain the implications of our low-noise analysis framework for the classical asymptotics.

Towards that goal, we will consider the scenarios where the sample size n is much larger than the dimension p. Analytically, we let δ go to infinity and calculate the expansions for AMSE in terms of large δ (similar to what we did in Chapter 3.2.2 for low noise). In this section, we write $\text{AMSE}(\lambda_{*,q}, q, \delta)$ for $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ to make it clear that the expansion is derived in terms of δ. Before getting to the results, we should clarify an important issue. Recall the definition of a converging sequence in Chapter 1.2. It is straightforward to confirm that the signal-to-noise ratio of each measurement is $\text{SNR} \propto \frac{\mathbb{E}|B|^2}{\delta \sigma_w^2}$. Hence if we take $\delta \to \infty$, SNR of each measurement will go to zero and this is inconsistent with the classical asymptotic setting where the SNR is in general assumed to be fixed. To fix this inconsistency, we will scale the noise term and consider a scaled linear model as follows,

$$y = X\beta + \frac{w}{\sqrt{\delta}}, \tag{3.1}$$

where $\{X, \beta, w\}$ is the converging sequence specified in Chapter 1.2. With the SNR remained a positive constant, this model is well aligned with the classical setting. Again for comparison purposes we start with the ordinary least squares estimate.

Lemma 3.3.1. *Consider the model* (3.1) *and OLS estimate* $\hat{\beta}(0, q)$*. Then as* $\delta \to \infty$*,*

$$\text{AMSE}(0, q, \delta) = \frac{\sigma_w^2}{\delta} + \frac{\sigma_w^2}{\delta^2} + o(\delta^{-2}).$$

Proof. This lemma is a simple application of Corollary 3.2.2. Under model (3.1), Corollary 3.2.2 shows that $\text{AMSE}(0, q, \delta) = \frac{\sigma_w^2}{\delta - 1}$. As $\delta \to \infty$, the expansion can be easily verified. □

We now discuss the bridge estimators with $q \in [1, 2]$.

Theorem 3.3.2. *Consider the model* (3.1)*. Suppose that* $\mathbb{P}(|B| > \mu) = 1$ *with* μ *being a positive constant and* $\mathbb{E}|B|^2 < \infty$*. Then for* $q \in [1, 2]$*, as* $\delta \to \infty$*,*

$$\text{AMSE}(\lambda_{*,q}, q, \delta) = \frac{\sigma_w^2}{\delta} + \frac{\sigma_w^2}{\delta^2} \cdot \frac{\mathbb{E}|B|^{2q-2} - (q-1)^2(\mathbb{E}|B|^{q-2})^2\sigma_w^2}{\mathbb{E}|B|^{2q-2}} + o(\delta^{-2}).$$

The proof can be found in Chapter 3.4.5. Since both Theorems 3.2.3 and 3.3.2 are concerned with coefficients that are bounded away from zero, we can compare their results. Again all of the LQLS have the same first dominant term. However, in the large sample regime, the second order term of LASSO is at the same order as that of other LQLS. Interestingly, the comparison of the constant in the second order term is consistent with that in the low noise case. Hence we obtain the same conclusions for two-point mixture distributions. For instance, bridge with $q \in (1, 2]$ outperforms OLS and $q = 2$ is optimal when all the mass is concentrated at one point. See Lemma 3.2.4 for more information on the comparison of C_q.

We now discuss the implications of Theorem 3.3.2 for classical asymptotics. In the classical setting where $n \to \infty$ and p is fixed, the performance of LQLS has been studied in Knight and Fu (2000). In particular the LQLS estimates were shown to have the regular $\sqrt{n}$ convergence. In our setting, we first let $n/p \to \delta$ and then $\delta \to \infty$. If we apply Theorem 2 in Knight and Fu (2000) to (3.1), a straightforward calculation for the asymptotic variance will give us the first dominant term in $\text{AMSE}(\lambda_{*,q}, q, \delta)$. In other words, the classical asymptotic result for LQLS in Knight and Fu (2000) only provides the "first-order" information regarding mean square error, and it is the same for all the values of $q \in [1, 2]$ under optimal tuning. The virtue of our asymptotic framework is to offer the second order term that can be used to evaluate and compare LQLS more accurately. The same can be derived when coefficients have mass around zero, as presented in the next theorem.

Theorem 3.3.3. *Consider the model* (3.1) *and assume* $\mathbb{E}|B|^2 < \infty$. *For any given* $q \in (1, 2)$, *suppose that* $\mathbb{P}(|B| \le t) = O(t^{2-q+\epsilon})$ *(as* $t \to 0$*) with* ϵ *being any positive constant, then as* $\delta \to \infty$,

$$\text{AMSE}(\lambda_{*,q}, q, \delta) = \frac{\sigma_w^2}{\delta} + \frac{\sigma_w^2}{\delta^2} \cdot \frac{\mathbb{E}|B|^{2q-2} - (q-1)^2(\mathbb{E}|B|^{q-2})^2\sigma_w^2}{\mathbb{E}|B|^{2q-2}} + o(\delta^{-2}),$$

for $q = 2$, *as* $\delta \to \infty$,

$$\text{AMSE}(\lambda_{*,q}, q, \delta) = \frac{\sigma_w^2}{\delta} + \frac{\sigma_w^2}{\delta^2} \cdot \frac{\mathbb{E}|B|^2 - \sigma_w^2}{\mathbb{E}|B|^2} + o(\delta^{-2}),$$

and for $q = 1$, *suppose* $\mathbb{P}(|B| \le t) = \Theta(t^\ell)$ *with* $0 < \ell < 1$, *then as* $\delta \to \infty$,

$$-|\Theta(\delta^{-\ell-1})| \cdot \big(\underbrace{\log\log\ldots\log}_{m \text{ times}}(\sqrt{\delta})\big)^\ell \lesssim \text{AMSE}(\lambda_{*,q}, q, \delta) - \frac{\sigma_w^2}{\delta} \lesssim -|\Theta(\delta^{-\ell-1})|,$$

where m *can be any natural number.*

The proof is presented in Chapter 3.4.6. Theorem 3.3.3 can be compared with Theorem 3.2.5. Again we see that the expansion for $q \in (1, 2]$ remains the same for more general coefficients, while the second order term of LASSO becomes order-wise smaller when coefficients put more mass around zero. For a given $q \in (1, 2]$, it is clear that LASSO outperforms this LQLS when $\mathbb{P}(|B| \le t) = \Theta(t^{2-q+\epsilon})$ with

$\epsilon \in (0, q-1)$. This implies that even in the case when n is much larger than p, if the underlying coefficient has many elements of small values, ℓ_1 regularization will improve the performance, which is characterized by a second order analysis that is not available from the $\sqrt{n}$ convergence result. Regrading the distributions with tail $e^{-\tau b^{q_0}}$, we see that the comparison among $q \in (1,2]$ in the low noise regime carries over.

The fact that regularization can improve the performance of the maximum likelihood estimate (i.e., OLS in the context of linear regression with Gaussian noise), seems to be contradictory with the classical results that imply MLE is asymptotically optimal under mild regularity conditions. However, note that the optimality of MLE is concerned with the asymptotic variance (equivalently the first order term) of the estimate. Our results show that many estimators share that first order term, while their actual performance might be different. Second dominant terms provide much more accurate information in these cases.

3.4 Proofs of the main results

3.4.1 Notations and Preliminaries

Throughout the proofs, B will be a random variable having the probability measure f_β that appears in the definition of the converging sequence, and Z will refer to a standard normal random variable. We will also use $\phi(\cdot)$ to denote the density function of Z and $F(b)$ to represent the cumulative distribution function of $|B|$. We further define the following useful notations:

$$R_q(\chi, \sigma) = \mathbb{E}(\eta_q(B/\sigma + Z; \chi) - B/\sigma)^2, \quad \chi_q^*(\sigma) = \arg\min_{\chi \geq 0} R_q(\chi, \sigma), \tag{3.2}$$

where B and Z are independent. Recall the proximal operator function $\eta_q(u;\chi)$. Since we will be using $\eta_q(u;\chi)$ extensively in the later proofs, we present some useful properties of $\eta_q(u;\chi)$ in the next lemma. Because $\eta_q(u;\chi)$ has explicit forms when

$q = 1, 2$, we focus on the case $1 < q < 2$. For notational simplicity we may use $\partial_i f(x_1, x_2, \ldots)$ to represent the partial derivative of f with respect to its ith argument.

Lemma 3.4.1. *For $q \in (1, 2)$, the function $\eta_q(u; \chi)$ satisfies the following properties.*

(i) $-\eta_q(u; \chi) = \eta_q(-u; \chi)$.

(ii) $u = \eta_q(u; \chi) + \chi q(q-1)\eta_q(u; \chi)\text{sign}(u)$.

(iii) $\alpha\eta_q(u; \chi) = \eta_q(\alpha u; \alpha^{2-q}\chi)$, *for* $\alpha > 0$.

(iv) $\frac{\partial \eta_q(u;\chi)}{\partial u} = \frac{1}{1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2}}$

(v) $\frac{\partial \eta_q(u;\chi)}{\partial \chi} = \frac{-q|\eta_q(u;\chi)|^{q-1}\text{sign}(u)}{1+\chi q(q-1)|\eta_q(u;\chi)|^{q-2}}$

(vi) *The function $\partial_2 \eta_q(u; \chi)$ is differentiable with respect to u.*

Proof. The proof has already been presented in Chapter 2.5.3. □

We next write down the Stein's lemma (Stein, 1981) that we will apply several times in the proofs.

Stein's lemma. *Suppose the function $f : \mathbb{R} \to \mathbb{R}$ is weakly differentiable and* $\mathbb{E}|f'(Z)| < \infty$, then

$$\mathbb{E}(Zf(Z)) = \mathbb{E}f'(Z).$$

3.4.2 Proof of Corollary 3.2.2

Since $\delta > 1$, $\hat{\beta}(0, q) = (X'X)^{-1}Xy$ is well defined with probability 1 for sufficiently large n. We first derive $\text{AMSE}(\lambda, 2, \sigma_w)$ for the ridge estimate $\hat{\beta}(\lambda, 2) = (X'X + \lambda I)^{-1}X'y$, and then obtain the AMSE for OLS by letting $\lambda \to 0$. According to Theorem 2.2.2, it is known that for given $\lambda > 0$,

$$\text{AMSE}(\lambda, 2, \sigma_w) = \delta(\sigma^2 - \sigma_w^2),$$

where σ is the solution to the following equation:

$$\sigma^2 = \sigma_w^2 + \frac{4\chi^2\mathbb{E}|B|^2 + \sigma^2}{\delta(1 + 2\chi)^2}, \quad \lambda = \chi - \frac{\chi}{\delta(1 + 2\chi)}.$$

After a few calculations we can obtain

$$\text{AMSE}(\lambda, 2, \sigma_w) = \frac{\delta(4\chi^2 \mathbb{E}|B|^2 + \sigma_w^2)}{\delta(1+2\chi)^2 - 1}, \tag{3.3}$$

with $\chi = \frac{1-\delta+2\lambda\delta+\sqrt{(\delta-1-2\lambda\delta)^2+8\lambda\delta^2}}{4\delta}$. Clearly $\text{AMSE}(\lambda, 2, \sigma_w) \to \frac{\sigma_w^2}{1-1/\delta}$ as $\lambda \to 0$. We now utilize that result to derive AMSE for OLS. According to the identity below

$$(X'X+\lambda I)^{-1} = (X'X)^{-1} - \lambda \underbrace{(X'X)^{-1}(I+\lambda(X'X)^{-1})^{-1}(X'X)^{-1}}_{H},$$

we have

$$\begin{aligned} & \frac{1}{p}\|\hat{\beta}(0,q)-\beta\|_2^2 - \frac{\sigma_w^2}{1-1/\delta} \\ = \; & \underbrace{\frac{1}{p}\|\hat{\beta}(\lambda,2)-\beta\|_2^2 - \frac{\sigma_w^2}{1-1/\delta}}_{J_1} + \underbrace{\frac{1}{p}\|\lambda HX'y\|_2^2}_{J_2} + \underbrace{\frac{2}{p}\langle \hat{\beta}(\lambda,2)-\beta, \lambda HX'y\rangle}_{J_3} \end{aligned} \tag{3.4}$$

Let $\sigma_{\min}(X)$ be the smallest non-zero singular values of X. It is not hard to confirm

$$\|HX'Y\|_2 \le \|HX'X\beta\|_2 + \|HX'w\|_2 \le \frac{\|\beta\|_2}{\lambda+\sigma_{\min}^2(X)} + \frac{\|w\|_2}{(\lambda+\sigma_{\min}^2(X))\sigma_{\min}^2(X)}.$$

Since $\sigma_{\min} \overset{a.s.}{\to} 1 - \frac{1}{\sqrt{\delta}} > 0$ (Bai and Yin, 1993) and β, w belong to the converging sequence defined in Chapter 1.2, we can conclude that $J_2 = O(\lambda^2), a.s.$. Moreover, we obtain from (3.3) that almost surely

$$J_1 = \frac{\delta(4\chi^2\mathbb{E}|B|^2+\sigma_w^2)}{\delta(1+2\chi)^2-1} - \frac{\sigma_w^2}{1-1/\delta}$$

The results on J_1, J_2 imply that $J_3 = O(\lambda), a.s.$. Further note that the term on the left hand side of (3.4) does not depend on λ. Therefore by letting $n \to \infty$ and then $\lambda \to 0$ on both sides of (3.4) finishes the proof.

3.4.3 Proof of Theorem 3.2.3

3.4.3.1 Roadmap

Since the proof has several long steps, we lay out the roadmap to help readers navigate through the details. According to Corollary 2.2.4, we know

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}), \tag{3.5}$$

where $\bar{\sigma}$ is the unique solution of

$$\bar{\sigma}^2 = \sigma_w^2 + \frac{\bar{\sigma}^2}{\delta} R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}). \tag{3.6}$$

Note from the above equation that $\bar{\sigma}$ is a function of σ_w. In the regime $\sigma_w \to 0$, we will show $\bar{\sigma} \to 0$. This fact combined with (3.5) tells us that in order to derive the second-order expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ as a function of σ_w, it is sufficient to characterize the convergence rate of $\bar{\sigma}$ as $\sigma_w \to 0$ and $R_q(\chi_q^*(\sigma), \sigma)$ as $\sigma \to 0$. For that purpose, we will first study the convergence rate of $\chi_q^*(\sigma)$ as $\sigma \to 0$, which will then enables us to obtain the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$. We then utilize that result and (3.6) to derive the rate of $\bar{\sigma}$ as $\sigma_w \to 0$. We give the proof for $1 < q \leq 2$ and $q = 1$ in Chapters 3.4.3.2 and 3.4.3.3, respectively.

3.4.3.2 Proof for the case $1 < q \leq 2$

Due to the explicit form of $\eta_2(u;\chi) = \frac{u}{1+2\chi}$, all the results for $q = 2$ in this section can be easily verified. We thus focus the proof on $1 < q < 2$.

Lemma 3.4.2. *Let $\chi_q^*(\sigma)$ be the optimal threshold value as defined in* (3.2). *Then $\chi_q^*(\sigma) \to 0$ as $\sigma \to 0$.*

Proof. The proof is essentially the same as the one for Lemma 2.5.30. Hence we do not repeat the arguments here. □

Lemma 3.4.3. *For $q \in (1, 2]$, suppose that $\mathbb{P}(|B| > \mu) = 1$ with μ being a positive constant and $\mathbb{E}|B|^2 < \infty$. Then as $\sigma \to 0$*

$$R_q(C\sigma^q, \sigma) = 1 + (C^2 q^2 \mathbb{E}|B|^{2q-2} - 2Cq(q-1)\mathbb{E}|B|^{q-2})\sigma^2 + o(\sigma^2),$$

where C is any fixed positive constant.

Proof. We aim to derive the convergence rate of $R_q(\chi, \sigma)$ when $\chi = C\sigma^q$. In this proof, we may write χ to denote $C\sigma^q$ for notational simplicity. According to Lemma

3.4.1 parts (ii)(iv) and Stein's lemma, we have the following formula for $R_q(\chi, \sigma)$:

$$R_q(\chi,\sigma)-1 = \mathbb{E}(\eta_q(B/\sigma+Z;\chi)-B/\sigma-Z)^2 + 2\mathbb{E}Z(\eta_q(B/\sigma+Z;\chi)-B/\sigma-Z)$$
$$= \underbrace{\chi^2 q^2 \mathbb{E}|\eta_q(B/\sigma+Z;\chi)|^{2q-2}}_{S_1} \underbrace{-2\chi q(q-1)\mathbb{E}\frac{|\eta_q(B/\sigma+Z;\chi)|^{q-2}}{1+\chi q(q-1)|\eta_q(B/\sigma+Z;\chi)|^{q-2}}}_{S_2}. \quad (3.7)$$

It is straightforward to confirm the following

$$\begin{aligned} \lim_{\sigma\to 0}\frac{S_1}{\sigma^2} &= \lim_{\sigma\to 0}\frac{\chi^2 q^2 \mathbb{E}|\eta_q(B/\sigma+Z;\chi)|^{2q-2}}{\sigma^2} \\ &= C^2q^2\lim_{\sigma\to 0}\mathbb{E}|\eta_q(B+\sigma Z;\chi\sigma^{2-q})|^{2q-2} = C^2q^2\mathbb{E}|B|^{2q-2}. \end{aligned} \quad (3.8)$$

The last equality is obtained by Dominated Convergence Theorem (DCT). The condition of DCT holds due to Lemma 3.4.1 part (ii). We now focus on analyzing S_2. We obtain

$$\begin{aligned} \frac{-S_2}{\sigma^2} &= 2C\sigma^{q-2}q(q-1)\mathbb{E}\frac{|\eta_q(B/\sigma+Z;\chi)|^{q-2}}{1+\chi q(q-1)|\eta_q(B/\sigma+Z;\chi)|^{q-2}} \\ &= 2C\sigma^{q-2}q(q-1)\mathbb{E}\frac{1}{|\eta_q(|B|/\sigma+Z;\chi)|^{2-q}+\chi q(q-1)} \\ &= \underbrace{2C\sigma^{q-2}q(q-1)\int_\mu^\infty\int_{-b/\sigma-\mu/(2\sigma)}^{-b/\sigma+\mu/(2\sigma)}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)}_{T_1} \end{aligned}$$
$$+\underbrace{2C\sigma^{q-2}q(q-1)\int_\mu^\infty\int_{\mathbb{R}\backslash[-b/\sigma-\mu/(2\sigma),-b/\sigma+\mu/(2\sigma)]}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)}_{T_2}.$$

We then consider T_1 and T_2 separately. For T_1, we have

$$\begin{aligned} T_1 &\leq 2C\sigma^{q-2}q(q-1)\int_\mu^\infty\int_{-\mu/\sigma-\mu/(2\sigma)}^{-\mu/\sigma+\mu/(2\sigma)}\frac{1}{\chi q(q-1)}\phi(z)dzdF(b) \\ &\leq 2\sigma^{-3}\mu\phi(\mu/(2\sigma)) \to 0, \text{ as } \sigma\to 0. \end{aligned} \quad (3.9)$$

Regarding T_2, DCT enables us to conclude

$$\begin{aligned} \lim_{\sigma\to 0}T_2 &= \lim_{\sigma\to 0}2C\sigma^{q-2}q(q-1)\mathbb{E}\frac{\mathbb{1}(Z\notin[-|B|/\sigma-\mu/(2\sigma),-|B|/\sigma+\mu/(2\sigma)])}{|\eta_q(|B|/\sigma+Z;\chi)|^{2-q}+\chi q(q-1)} \\ &= \lim_{\sigma\to 0}2Cq(q-1)\mathbb{E}\frac{\mathbb{1}(Z\notin[-|B|/\sigma-\mu/(2\sigma),-|B|/\sigma+\mu/(2\sigma)])}{|\eta_q(|B|+\sigma Z;\chi\sigma^{2-q})|^{2-q}+C\sigma^2q(q-1)} \\ &= 2Cq(q-1)\mathbb{E}|B|^{q-2}. \end{aligned} \quad (3.10)$$

Note that DCT works here because for small enough σ, Lemma 3.4.1 parts (iv)(v) implies

$$\frac{\mathbb{1}(Z \notin [-|B|/\sigma - \mu/(2\sigma), -|B|/\sigma + \mu/(2\sigma)])}{|\eta_q(|B| + \sigma Z; \chi\sigma^{2-q})|^{2-q} + C\sigma^2 q(q-1)} \le \frac{1}{|\eta_q(\mu/2; \chi\sigma^{2-q})|^{2-q}} \le \frac{1}{|\eta_q(\mu/2; 1)|^{2-q}}.$$

Combining (3.7), (3.8), (3.9) and (3.10) together completes the proof. □

Lemma 3.4.3 shows that by choosing an appropriate χ for σ small enough, $R_q(\chi, \sigma)$ is less than 1. This result will be used to show that $\chi_q^*(\sigma)$ cannot converge to zero too fast. We then utilize this fact to derive the exact convergence rate of $\chi_q^*(\sigma)$. This is done in the next lemma.

Lemma 3.4.4. *Suppose that $\mathbb{P}(|B| > \mu) = 1$ with μ being a positive constant and $\mathbb{E}|B|^2 < \infty$, then for $q \in (1, 2]$ we have as $\sigma \to 0$*

$$\chi_q^*(\sigma) = \frac{(q-1)\mathbb{E}|B|^{q-2}}{q\mathbb{E}|B|^{2q-2}}\sigma^q + o(\sigma^q),$$

$$R_q(\chi_q^*(\sigma), \sigma) = 1 - \frac{(q-1)^2(\mathbb{E}|B|^{q-2})^2}{\mathbb{E}|B|^{2q-2}}\sigma^2 + o(\sigma^2).$$

Proof. Choosing $\chi = \frac{(q-1)\mathbb{E}|B|^{q-2}}{q\mathbb{E}|B|^{2q-2}} \cdot \sigma^q$ in Lemma 3.4.3, we have

$$\lim_{\sigma \to 0} \frac{R_q(\chi, \sigma) - 1}{\sigma^2} = -\frac{(q-1)^2(\mathbb{E}|B|^{q-2})^2}{\mathbb{E}|B|^{2q-2}} < 0. \tag{3.11}$$

That means for sufficiently small σ

$$R_q(\chi_q^*(\sigma), \sigma) \le R_q(\chi, \sigma) < 1 = R_q(0, \sigma).$$

Hence we can conclude that $\chi_q^*(\sigma) > 0$ when σ is small enough. Moreover, by a slight change of arguments in the proof of Lemma 3.4.3 summarized below:

1. the fact $\chi\sigma^{2-q} = o(1)$ used several times in Lemma 3.4.3 still holds here
2. $\chi\sigma^{2-q} = o(1)$ and $\chi = o(\sigma^q)$ are sufficient to have $S_1 = o(\sigma^2)$
3. bounding the term T_1 in (3.9) does not depend on χ
4. $\chi\sigma^{2-q} = o(1)$ and $\chi = o(\sigma^q)$ are sufficient to obtain $T_2 = o(1)$

we can show

$$\lim_{\sigma \to 0} \frac{R_q(\chi, \sigma) - 1}{\sigma^2} = 0, \tag{3.12}$$

for $\chi = O(\exp(-c/\sigma))$ with any fixed positive constant c. This implies that $\lim_{\sigma \to 0} \chi_q^*(\sigma) \cdot e^{c/\sigma} = +\infty$ for any $c > 0$. Otherwise there exists a sequence $\sigma_n \to 0$ such that $\chi_q(\sigma_n) e^{c/\sigma_n} = O(1)$. This result combined with (3.11) and (3.12) contradicts with the fact that $\chi = \chi_q^*(\sigma)$ is the minimizer of $R_q(\chi, \sigma)$. We will use the two aforementioned properties of $\chi_q^*(\sigma)$ we have showed so far in the following proof. For notational simplicity, in the rest of the proof we may use χ to denote $\chi_q^*(\sigma)$ whenever no confusion is caused. Firstly since $\chi_q^*(\sigma)$ is a non-zero finite value, it is a solution of the first order optimality condition $\frac{\partial R_q(\chi, \sigma)}{\partial \chi} = 0$, which can be further written out as

$$\begin{aligned}
0 &= \mathbb{E}((\eta_q(B/\sigma + Z; \chi) - B/\sigma)\partial_2 \eta_q(B/\sigma + Z; \chi)) \\
&\overset{(a)}{=} \mathbb{E}\frac{-(\eta_q(B/\sigma + Z; \chi) - B/\sigma - Z) q |\eta_q(B/\sigma + Z; \chi)|^{q-1} \text{sign}(B/\sigma + Z)}{1 + \chi q(q-1)|\eta_q(B/\sigma + Z; \chi)|^{q-2}} \\
&\quad + \mathbb{E}(Z \partial_2 \eta_q(B/\sigma + Z; \chi)) \\
&\overset{(b)}{=} \underbrace{\chi\, \mathbb{E}\frac{q^2 |\eta_q(B/\sigma + Z; \chi)|^{2q-2}}{1 + \chi q(q-1)|\eta_q(B/\sigma + Z; \chi)|^{q-2}}}_{U_1} - \underbrace{\mathbb{E}\frac{q(q-1)|\eta_q(B/\sigma + Z; \chi)|^{4-2q}}{(|\eta_q(B/\sigma + Z; \chi)|^{2-q} + \chi q(q-1))^3}}_{U_2} \\
&\quad \underbrace{- \chi\, \mathbb{E}\frac{q^2(q-1)|\eta_q(B/\sigma + Z; \chi)|^{2-q}}{(|\eta_q(B/\sigma + Z; \chi)|^{2-q} + \chi q(q-1))^3}}_{U_3}.
\end{aligned} \tag{3.13}$$

We have used Lemma 3.4.1 part (v) to derive (a). To obtain (b), we have used the following steps:

1. We used Lemma 3.4.1 part (ii) to conclude that

$$\eta_q(B/\sigma + Z; \chi) - B/\sigma - Z = -\chi q |\eta_q(B/\sigma + Z; \chi)|^{q-1} \text{sign}(B/\sigma + Z).$$

2. We used the expression we derived in Lemma 3.4.1 part (v) for $\partial_2 \eta_q(B/\sigma + Z; \chi)$ and then employed Stein's lemma to simplify $\mathbb{E}(Z \partial_2 \eta_q(B/\sigma + Z; \chi))$. Note that according to Lemma 3.4.1 part (vi), $\partial_2 \eta_q(B/\sigma + Z; \chi)$ is differentiable with respect to its first argument and hence Stein's lemma can be applied.

We now evaluate the three terms U_1, U_2 and U_3 individually. Our goal is to show the following:

(i) $\lim_{\sigma\to 0}\sigma^{2q-2}U_1 = q^2\mathbb{E}|B|^{2q-2}$.

(ii) $\lim_{\sigma\to 0}\sigma^{q-2}U_2 = q(q-1)\mathbb{E}|B|^{q-2}$.

(iii) $\lim_{\sigma\to 0}\sigma^{2q-4}U_3 = q^2(q-1)\mathbb{E}|B|^{2q-4}$.

For the term U_1, we can apply Dominated Convergence Theorem (DCT)

$$\lim_{\sigma\to 0}\sigma^{2q-2}U_1 = \mathbb{E}\lim_{\sigma\to 0}\frac{q^2|\eta_q(B+\sigma Z;\chi\sigma^{2-q})|^{2q-2}}{1+\chi\sigma^{2-q}q(q-1)|\eta_q(B+\sigma Z;\chi\sigma^{2-q})|^{q-2}} = q^2\mathbb{E}|B|^{2q-2}.$$

We now derive the convergence rate of U_2. We have

$$\begin{aligned} U_2 &= \int_{\mu}^{\infty}\int_{-\infty}^{\infty}\frac{q(q-1)|\eta_q(b/\sigma+z;\chi)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1))^3}\phi(z)dzdF(b) \\ &= \underbrace{\int_{\mu}^{\infty}\int_{-\frac{b}{\sigma}-\frac{\mu}{2\sigma}}^{-\frac{b}{\sigma}+\frac{\mu}{2\sigma}}\frac{q(q-1)|\eta_q(b/\sigma+z;\chi)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1))^3}\phi(z)dzdF(b)}_{U_{21}} \\ &+ \underbrace{\int_{\mu}^{\infty}\int_{z\notin[-\frac{b}{\sigma}-\frac{\mu}{2\sigma},-\frac{b}{\sigma}+\frac{\mu}{2\sigma}]}\frac{q(q-1)|\eta_q(b/\sigma+z;\chi)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1))^3}\phi(z)dzdF(b)}_{U_{22}}. \end{aligned} \tag{3.14}$$

First note that

$$\begin{aligned} \sigma^{q-2}U_{21} &\le \sigma^{q-2}\int_{\mu}^{\infty}\int_{-\frac{b}{\sigma}-\frac{\mu}{2\sigma}}^{-\frac{b}{\sigma}+\frac{\mu}{2\sigma}}\frac{q(q-1)(\mu/(2\sigma))^{4-2q}}{(\chi q(q-1))^3}\phi(z)dzdF(b) \\ &\le \frac{\mu^{5-2q}\phi(\mu/(2\sigma))}{\sigma^{7-3q}2^{4-2q}\chi^3q^2(q-1)^2}\to 0,\ \text{as } \sigma\to 0, \end{aligned} \tag{3.15}$$

where the last step is due to the fact that $\lim_{\sigma\to 0}\chi e^{c/\sigma} = +\infty$. To evaluate U_{22} we first derive the following bounds for small enough σ

$$\begin{aligned} &\frac{\mathbb{1}(z\notin[-\frac{b}{\sigma}-\frac{\mu}{2\sigma},-\frac{b}{\sigma}+\frac{\mu}{2\sigma}])\cdot q(q-1)|\eta_q(b+\sigma z;\chi\sigma^{2-q})|^{4-2q}}{(|\eta_q(b+\sigma z;\chi\sigma^{2-q})|^{2-q}+\chi\sigma^{2-q}q(q-1))^3} \\ &\le \frac{q(q-1)}{|\eta_q(\mu/2;\chi\sigma^{2-q})|^{2-q}} \le \frac{q(q-1)}{|\eta_q(\mu/2;1)|^{2-q}}. \end{aligned}$$

Hence we are able to apply DCT to obtain

$$\lim_{\sigma\to 0} \sigma^{q-2} U_{22} = q(q-1)\mathbb{E}|B|^{q-2}. \tag{3.16}$$

Combining (3.14), (3.15), and (3.16) proves the result (ii). We can use similar arguments to show result (iii). Finally, we utilize the convergence results for U_1, U_2, U_3 and Equation (3.13) to derive

$$\lim_{\sigma\to 0} \frac{\chi}{\sigma^q} = \lim_{\sigma\to 0} \frac{\lim_{\sigma\to 0} \sigma^{q-2} U_2}{\lim_{\sigma\to 0} \sigma^{2q-2} U_2 - \lim_{\sigma\to 0} \sigma^{2q-2} U_3} = \frac{(q-1)\mathbb{E}|B|^{q-2}}{q\mathbb{E}|B|^{2q-2}}.$$

Now since we know the exact convergence order of $\chi_q^*(\sigma)$, (3.11) shows the exact order of $R_q(\chi_q^*(\sigma), \sigma)$. □

We are in position to derive the second-order expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ as $\sigma_w \to 0$ for $q \in (1, 2]$. According to Equation (3.6) and the fact that $\chi = \chi_q^*(\bar{\sigma})$ minimizes $R_q(\chi, \bar{\sigma})$, it is clear that $\delta(\bar{\sigma}^2 - \sigma_w^2) \le \bar{\sigma}^2 R_q(0, \bar{\sigma}) = \bar{\sigma}^2$, which combined with the condition $\delta > 1$ implies $\bar{\sigma} \to 0$ as $\sigma_w \to 0$. This result further enables us to conclude from (3.6):

$$\lim_{\sigma_w \to 0} \frac{\bar{\sigma}^2}{\sigma_w^2} = \frac{\delta}{\delta - 1}, \tag{3.17}$$

where we have used $R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) \to 1$ from Lemma 3.4.4. We finally utilize Lemma 3.4.4, Equations (3.5), (3.6) and (3.17) to derive the expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ in the following way:

$$\begin{aligned}
&\sigma_w^{-4}\Big(\text{AMSE}(\lambda_{*,q}, q, \sigma_w) - \frac{\sigma_w^2}{1 - 1/\delta}\Big) \\
&= \sigma_w^{-4}\Big(\bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - \frac{\delta}{\delta - 1}(\bar{\sigma}^2 - \frac{1}{\delta}\bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}))\Big) \\
&= \frac{\delta}{\delta - 1} \cdot \frac{\bar{\sigma}^4}{\sigma_w^4} \cdot \frac{R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1}{\bar{\sigma}^2} \to \frac{-\delta^3 (q-1)^2 (\mathbb{E}|B|^{q-2})^2}{(\delta - 1)^3 \mathbb{E}|B|^{2q-2}}.
\end{aligned}$$

This completes the proof of Theorem 3.2.3 for $q \in (1, 2]$.

3.4.3.3 Proof for the case $q = 1$

Lemma 3.4.5. *Suppose that $\mathbb{P}(|B| > \mu) = 1$ with μ being a positive constant and $\mathbb{E}|B|^2 < \infty$, then for $q = 1$ as $\sigma \to 0$*

$$\chi_q^*(\sigma) = O(\phi(\mu/\sigma)), \quad R_q(\chi_q^*(\sigma), \sigma) - 1 = O(\phi^2(\mu/\sigma)).$$

Proof. We first claim that $\chi_q^*(\sigma) \to 0$ as $\sigma \to 0$. Otherwise, there exists a sequence $\sigma_n \to 0$ such that $\chi_q^*(\sigma_n) \to C > 0$ as $n \to \infty$. And the limit C is finite. Suppose this is not true, then since $\eta_1(u; \chi) = \text{sign}(u)(|u| - \chi)_+$ we can apply Fatou's lemma to conclude

$$\liminf_{n\to\infty} R_q(\chi_q^*(\sigma_n), \sigma_n) \geq \mathbb{E} \liminf_{n\to\infty} (\eta_1(B/\sigma_n + Z; \chi_q^*(\sigma_n)) - B/\sigma_n)^2 = +\infty,$$

contradicting with the fact $R_q(\chi_q^*(\sigma_n), \sigma_n) \leq R_q(0, \sigma_n) = 1$. We now calculate the following limit:

$$\begin{aligned}&\lim_{n\to\infty} R_q(\chi_q^*(\sigma_n), \sigma_n) = \lim_{n\to\infty} \mathbb{E}(\eta_1(B/\sigma_n + Z; \chi_q^*(\sigma_n)) - B/\sigma_n - Z)^2 \\ &+2 \lim_{n\to\infty} \mathbb{E} Z(\eta_1(B/\sigma_n + Z; \chi_q^*(\sigma_n)) - B/\sigma_n - Z) + 1 = C^2 + 1.\end{aligned}$$

The last step is due to Dominated Convergence Theorem (DCT). The condition of DCT can be verified based on the fact $|u - \eta_1(u; \chi)| \leq \chi$. We can also choose a positive constant $\tilde{C}$ smaller than C and use similar argument to obtain $\lim_{n\to\infty} R_q(\tilde{C}, \sigma_n) = \tilde{C}^2 + 1$. That means $R_q(\tilde{C}, \sigma_n) < R_q(\chi_q^*(\sigma_n), \sigma_n)$ when n is large enough. This is contradicting with the fact $\chi = \chi_q^*(\sigma_n)$ minimizes $R_q(\chi, \sigma_n)$.

We next derive the following bounds:

$$\begin{aligned} R_q(\chi, \sigma) - 1 &= \mathbb{E}(\eta_1(B/\sigma + Z; \chi) - B/\sigma - Z)^2 + 2\mathbb{E}(Z(\eta_1(B/\sigma + Z; \chi) - B/\sigma - Z)) \\ &\overset{(a)}{=} \mathbb{E}(\eta_1(B/\sigma + Z; \chi) - B/\sigma - Z)^2 + 2\mathbb{E}(\partial_1 \eta_1(B/\sigma + Z; \chi) - 1) \\ &\overset{(b)}{\leq} \chi^2 - 2\mathbb{E} \int_{-B/\sigma - \chi}^{-B/\sigma + \chi} \phi(z) dz \overset{(c)}{=} \chi^2 - 4\chi \mathbb{E}\phi(-B/\sigma + \alpha\chi). \end{aligned}$$

To obtain (a) we used Stein's lemma; note that $\eta_1(u; \chi)$ is a weakly differentiable function of u. Inequality (b) holds since $|\eta_1(u; \chi) - u| \leq \chi$. Equality (c) is the result of the mean value theorem and hence $|\alpha| \leq 1$ is dependent on B. From the above inequality, it is straightforward to verify that if we choose $\chi = 3e^{-1}\mathbb{E}\phi(\sqrt{2}B/\sigma)$, then

$$R_q(\chi_q^*(\sigma), \sigma) \leq R_q(\chi, \sigma) < 1 = R_q(0, \sigma), \tag{3.18}$$

for small enough σ. This means the optimal threshold $\chi_q^*(\sigma)$ is a non-zero finite value. Hence it is a solution to $\frac{\partial R_q(\chi_q^*(\sigma),\sigma)}{\partial \chi} = 0$, which further implies (from now on we use χ^* to represent $\chi_q^*(\sigma)$ for simplicity):

$$\begin{aligned} \chi^* &= \frac{\mathbb{E}\phi(\chi^* - B/\sigma) + \mathbb{E}\phi(\chi^* + B/\sigma)}{\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*)} \leq \frac{2\mathbb{E}\phi(\chi^* - |B|/\sigma)}{\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*)} \\ &\leq \frac{2\phi(\chi^* - \mu/\sigma)}{\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*)}, \end{aligned} \tag{3.19}$$

where the last inequality holds for small values of σ due to the condition $\mathbb{P}(|B| > \mu) = 1$. Since $\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*) \to 1$, as $\sigma \to 0$ and $\phi(\chi^* - \mu/\sigma) \leq \phi(\mu/(\sqrt{2}\sigma))e^{(\chi^*)^2/2}$, from (3.19) we can first conclude $\chi^* = o(\sigma)$, which in turn (use (3.19) again) implies $\chi^* = O(\phi(\mu/\sigma))$.

We now turn to analyzing $R_q(\chi^*, \sigma)$:

$$\begin{aligned} R_q(\chi^*, \sigma) - 1 &= \mathbb{E}(\eta_1(B/\sigma + Z; \chi^*) - B/\sigma - Z)^2 + 2\mathbb{E}(\partial_1\eta_1(B/\sigma + Z; \chi^*) - 1) \\ &\geq -2\mathbb{E}\mathbb{1}(|B/\sigma + Z| \leq \chi^*) \geq -2\int_{-\mu/\sigma-\chi^*}^{-\mu/\sigma+\chi^*} \phi(z)dz \geq -4\chi^*\phi(\mu/\sigma - \chi^*) \\ &\overset{(d)}{\geq} \frac{-8\phi^2(\chi^* - \mu/\sigma)}{\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*)} \overset{(e)}{\sim} -8\phi^2(\mu/\sigma), \end{aligned}$$

where (d) is due to (3.19) and (e) holds because $\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*) \to 1$ and $\chi^* = o(\sigma)$. This result combined with $R_q(\chi^*, \sigma) - 1 < 0$ from (3.18) finishes the proof. □

We are in position to derive the expansion of $\text{AMSE}(\lambda_{*,1}, 1, \sigma_w)$. Similarly as in the proof for $q \in (1, 2]$, we can use Lemma 3.4.5 to derive (3.17) for $q = 1$. Then we apply Lemma 3.4.5 again to obtain

$$\begin{aligned} &\text{AMSE}(\lambda_{*,1}, 1, \sigma_w) - \frac{\delta}{\delta - 1}\sigma_w^2 = \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - \frac{\delta}{\delta - 1}(\bar{\sigma}^2 - \bar{\sigma}^2 R_q(\sigma_q^*(\bar{\sigma}), \bar{\sigma})/\delta) \\ &= \frac{\delta\bar{\sigma}^2(R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1)}{\delta - 1} = o(\exp(-\bar{\mu}^2/\bar{\sigma}^2)) = o(\exp(-\tilde{\mu}^2(\delta - 1)/(\delta\sigma_w^2))), \end{aligned}$$

where $0 < \tilde{\mu} < \bar{\mu} < \mu$. This closes the proof.

3.4.4 Proof of Theorem 3.2.5

Similar to the proof of Theorem 3.2.3, we consider two cases, i.e. $1 < q \leq 2$ and $q = 1$, and prove them separately. We will follow closely the roadmap illustrated in Chapter 3.4.3.1.

3.4.4.1 Proof for the case $1 < q < 2$

Again all the results in this section can be proved easily for $q = 2$. We will only consider $1 < q < 2$. Before we start the proof of our main result, we mention a simple lemma that will be used multiple times in our proof.

Lemma 3.4.6. *Let $T(\sigma)$ and $\chi(\sigma)$ be two nonnegative sequences with the property: $\chi(\sigma)T^{q-2}(\sigma) \to 0$, as $\sigma \to 0$. Then,*

$$\lim_{\sigma\to 0} \frac{\eta_q(T(\sigma), \chi(\sigma))}{T(\sigma)} = 1.$$

Proof. The proof is a simple application of scale invariance property of η_q, i.e, Lemma 3.4.1 part (iii). We have

$$\lim_{\sigma\to 0} \frac{\eta_q(T(\sigma), \chi(\sigma))}{T(\sigma)} = \lim_{\sigma\to 0} \eta_q(1; \chi(\sigma)T^{q-2}(\sigma)) = 1,$$

where the last step is the result of Lemma 3.4.1 part (ii). □

Our first goal is to show that when $\chi = C\sigma^q$, then $\lim_{\sigma\to 0} \frac{R_q(\chi,\sigma)-1}{\sigma^2}$ is a negative constant by choosing an appropriate C. However, since this proof is long, we break it to several steps. These steps are summarized in Lemmas 3.4.7, 3.4.8, and 3.4.9. Then in Lemma 3.4.10 we employ these three results to show that if $\chi = C\sigma^q$, then

$$\lim_{\sigma\to 0} \frac{R(\chi, \sigma) - 1}{\sigma^2} = C^2 q^2 \mathbb{E}|B|^{2q-2} - 2Cq(q-1)\mathbb{E}|B|^{q-2}.$$

Lemma 3.4.7. *For any given $q \in (1,2)$, suppose that $\mathbb{P}(|B| < t) = O(t^{2-q+\epsilon})$ (as $t \to 0$) with ϵ being any positive constant, $\mathbb{E}|B|^2 < \infty$ and $\chi = C\sigma^q$, where $C > 0$ is a fixed number. Then we have*

$$\sigma^{q-2}\int_0^\infty \int_{\frac{-b}{\sigma}-\alpha}^{\frac{-b}{\sigma}+\alpha} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b) \to 0,$$

as $\sigma \to 0$. Note that α is an arbitrary positive constant.

Proof. The main idea of the proof is to break this integral into several pieces and prove that each piece converges to zero. Throughout the proof, we will choose ϵ small enough to be in $(0, q-1)$. Based on the value of q, we consider the following intervals. First find the unique non-negative integer of m^* such that

$$q \in [2-(\epsilon/(\epsilon+q-1))^{\frac{1}{m^*+1}}, 2-(\epsilon/(\epsilon+q-1))^{\frac{1}{m^*}}).$$

Denote $\mathcal{S}_m^n(l) = l^m + l^{m+1} + \cdots + l^n (m \le n)$. Now we define the following intervals:

$$\begin{aligned}
\mathcal{I}_{-1} &= \left[-\frac{b}{\sigma} - \frac{\sigma^{q-\epsilon}}{\log(\frac{1}{\sigma})}, -\frac{b}{\sigma} + \frac{\sigma^{q-\epsilon}}{\log(\frac{1}{\sigma})}\right], \\
\mathcal{I}_i &= \left[-\frac{b}{\sigma} - \frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^i - \frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^i(2-q)}}, -\frac{b}{\sigma} + \frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^i - \frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^i(2-q)}}\right], \quad 0 \le i \le m^*, \\
\mathcal{I}_{m^*+1} &= \left[-\frac{b}{\sigma} - \frac{1}{(\log(1/\sigma))^{\mathcal{S}_0^{m^*+1}(2-q)}}, -\frac{b}{\sigma} + \frac{1}{(\log(1/\sigma))^{\mathcal{S}_0^{m^*+1}(2-q)}}\right], \\
\mathcal{I}_{m^*+2} &= \left[\frac{-b}{\sigma} - \alpha, -\frac{b}{\sigma} + \alpha\right]. \qquad (3.20)
\end{aligned}$$

We see that for small enough σ, these intervals are nested: $\mathcal{I}_{-1} \subset \mathcal{I}_0 \subset \mathcal{I}_1 \subset \ldots \subset \mathcal{I}_{m^*+2}$. Further define

$$\begin{aligned}
P_{-1} &= \sigma^{q-2}\int_0^\infty \int_{\mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b), \\
P_i &= \sigma^{q-2}\int_0^\infty \int_{\mathcal{I}_i \backslash \mathcal{I}_{i-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b), \quad 0 \le i \le m^*+2.
\end{aligned}$$

Using these notations we have

$$\sigma^{q-2}\int_0^\infty \int_{\frac{-b}{\sigma}-\alpha}^{\frac{-b}{\sigma}+\alpha} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b) = \sum_{i=-1}^{m^*+2} P_i. \qquad (3.21)$$

Our goal is to show that $P_i \to 0$ as $\sigma \to 0$. Since these intervals have different forms, we consider five different cases (i) $i = -1$, (ii) $i = 0$, (iii) $1 \le i \le m^*$, (iv) $i = m^* + 1$, and (v) $i = m^* + 2$ and for each case we show that $P_i \to 0$. Let $|\mathcal{I}|$ denote the Lebesgue measure of an interval $\mathcal{I}$. For the first term, we have for a positive constant $\tilde{C}_{-1}$,

$$
\begin{aligned}
P_{-1} &\le \sigma^{q-2} \int_0^\infty \int_{\mathcal{I}_{-1}} \frac{1}{\chi q(q-1)} \phi(z) dz dF(b) \le \sigma^{q-2} \int_0^{\tilde{C}_{-1}\sigma\sqrt{\log(1/\sigma)}} \int_{\mathcal{I}_{-1}} \frac{1}{\chi q(q-1)} \phi(z) dz dF(b) \\
&\quad + \sigma^{q-2} \int_{\tilde{C}_{-1}\sigma\sqrt{\log(1/\sigma)}}^\infty \int_{\mathcal{I}_{-1}} \frac{1}{\chi q(q-1)} \phi(z) dz dF(b) \\
&\le \frac{\sigma^{q-2}\phi(0)|\mathcal{I}_{-1}|\mathbb{P}(|B| \le \tilde{C}_{-1}\sigma\sqrt{\log(1/\sigma)})}{\chi q(q-1)} + \frac{\sigma^{q-2}\phi(\tilde{C}_{-1}\sqrt{\log(1/\sigma)} - \frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})|\mathcal{I}_{-1}|}{\chi q(q-1)} \\
&\le O(1)\frac{\sigma^{q-\epsilon-2}\mathbb{P}(|B| \le \tilde{C}_{-1}\sigma\sqrt{\log(1/\sigma)})}{\log(1/\sigma)} + O(1)\frac{\sigma^{q-\epsilon-2}\phi(\frac{\tilde{C}_{-1}}{2}\sqrt{\log(1/\sigma)})}{\log(1/\sigma)} \\
&\le O(1)(\log(1/\sigma))^{\frac{-q+\epsilon}{2}} + O(1)\frac{\sigma^{q-\epsilon-2+\tilde{C}_{-1}^2/8}}{\log(1/\sigma)} \to 0, \qquad (3.22)
\end{aligned}
$$

where we have used the condition $\mathbb{P}(|B| < t) = O(t^{2-q+\epsilon})$ to obtain the last inequality and the last statement holds by choosing $\tilde{C}_{-1}$ large enough. We next analyze the term P_0. For a constant $\tilde{C}_0 > 0$ we have

$$
\begin{aligned}
P_0 &\le \sigma^{q-2} \int_0^\infty \int_{\mathcal{I}_0 \backslash \mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q}} \phi(z) dz dF(b) \\
&= \sigma^{q-2} \int_0^{\tilde{C}_0\sigma\sqrt{\log(1/\sigma)}} \int_{\mathcal{I}_0 \backslash \mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q}} \phi(z) dz dF(b) \\
&\quad + \sigma^{q-2} \int_{\tilde{C}_0\sigma\sqrt{\log(1/\sigma)}}^\infty \int_{\mathcal{I}_0 \backslash \mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q}} \phi(z) dz dF(b) \\
&\le \frac{\sigma^{q-2}\phi(0)|\mathcal{I}_0|\mathbb{P}(|B| < \tilde{C}_0\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)}; \chi\right)} + \frac{\sigma^{q-2}\phi\left(\tilde{C}_0\sqrt{\log(1/\sigma)} - \frac{\sigma}{\log(1/\sigma)}\right)|\mathcal{I}_0|}{\eta_q^{2-q}\left(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)}; \chi\right)}. \qquad (3.23)
\end{aligned}
$$

We have used the fact that $|b/\sigma + z| \ge \frac{\sigma^{q-\epsilon}}{\log(1/\sigma)}$ for $z \notin \mathcal{I}_{-1}$ in the last step. Note that according to Lemma 3.4.6, since $(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})^{q-2}\chi \propto \sigma^{q^2-(1+\epsilon)q+2\epsilon}(\log(1/\sigma))^{2-q} \to 0$, we obtain

$$
\lim_{\sigma \to 0} \frac{\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)}}{\eta_q(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)}; \chi)} = 1.
$$

With the above result, it is clear that the second term of the upper bound in (3.23) vanishes if choosing sufficiently large $\tilde{C}_0$. Regarding the first term we know

$$\frac{\sigma^{q-2}|\mathcal{I}_0|\mathbb{P}(|B|<\tilde{C}_0\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\big(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)};\chi\big)} \propto \sigma^{q^2-(\epsilon+2)q+3\epsilon+1}(\log(1/\sigma))^{\frac{\epsilon+4-3q}{2}} = o(1).$$

Now we consider an arbitrary $1 \leq i \leq m^*$ and show that $P_i \to 0$. Similarly as bounding P_0 we can have

$$\begin{aligned} P_i &\leq \sigma^{q-2}\int_0^{\tilde{C}_i\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{I}_i\backslash\mathcal{I}_{i-1}}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}}\phi(z)dzdF(b) \\ &\quad +\sigma^{q-2}\int_{\tilde{C}_i\sigma\sqrt{\log(1/\sigma)}}^{\infty}\int_{\mathcal{I}_i\backslash\mathcal{I}_{i-1}}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}}\phi(z)dzdF(b) \\ &\leq \frac{\sigma^{q-2}\phi(0)|\mathcal{I}_i|\mathbb{P}(|B|<\tilde{C}_i\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i-1}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i-1}(2-q)}};\chi\right)} \\ &\quad +\frac{\sigma^{q-2}\phi\left(\tilde{C}_i\sqrt{\log(1/\sigma)}-\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i}(2-q)}}\right)|\mathcal{I}_i|}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i-1}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i-1}(2-q)}};\chi\right)}. \end{aligned} \tag{3.24}$$

We then use Lemma 3.4.6 to conclude for $i \geq 1$

$$\lim_{\sigma\to 0}\frac{\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i}-\frac{\epsilon(2-q)}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_1^{i}(2-q)}}}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i-1}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i-1}(2-q)}};\chi\right)} = 1. \tag{3.25}$$

The condition of Lemma 3.4.6 can be verified in the following:

$$(\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i-1}-\frac{\epsilon}{q-1}}(\log(1/\sigma))^{-\mathcal{S}_0^{i-1}(2-q)})^{q-2}\chi \propto \sigma^{-\frac{\epsilon+q-1}{q-1}(2-q)^i+\frac{2-q}{q-1}\epsilon+q}(\log(1/\sigma))^{\mathcal{S}_1^{i}(2-q)} = o(1),$$

where the last step is due to the fact that

$$-\frac{\epsilon+q-1}{q-1}(2-q)^i+\frac{2-q}{q-1}\epsilon+q \geq -\frac{\epsilon+q-1}{q-1}(2-q)+\frac{2-q}{q-1}\epsilon+q = 2q-2 > 0.$$

Using the result (3.25), it is straightforward to confirm that if $\tilde{C}_i$ is chosen large enough, the second term in (3.24) goes to zero. For the first term,

$$\lim_{\sigma\to 0} \frac{\sigma^{q-2}\phi(0)|\mathcal{I}_i|\mathbb{P}(|B|<\tilde{C}_i\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i-1}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i-1}(2-q)}};\chi\right)}$$

$$\overset{(a)}{=} O(1)\cdot\lim_{\sigma\to 0}\frac{\sigma^{q-2}\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{i}(2-q)}}\sigma^{2-q+\epsilon}(\log(1/\sigma))^{\frac{2-q+\epsilon}{2}}}{\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{i}-\frac{\epsilon(2-q)}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_1^{i}(2-q)}}} = O(1)\cdot\lim_{\sigma\to 0}(\log(1/\sigma))^{\frac{-q+\epsilon}{2}} = 0,$$

where we have used (3.25) to obtain (a). So far we have showed $\lim_{\sigma\to 0}\sum_{i=-1}^{m^*} P_i = 0$. Our next step is to prove that $P_{m^*+1}\to 0$.

$$\begin{aligned} P_{m^*+1} &\leq \sigma^{q-2}\int_0^{\tilde{C}_{m^*+1}\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{I}_{m^*+1}\backslash\mathcal{I}_{m^*}}\frac{\phi(z)}{|\eta_q(b/\sigma+z;\chi)|^{2-q}}dzdF(b) \\ &\quad +\sigma^{q-2}\int_{\tilde{C}_{m^*+1}\sigma\sqrt{\log(1/\sigma)}}^{\infty}\int_{\mathcal{I}_{m^*+1}\backslash\mathcal{I}_{m^*}}\frac{\phi(z)}{|\eta_q(b/\sigma+z;\chi)|^{2-q}}dzdF(b) \\ &\leq \frac{\sigma^{q-2}\phi(0)|\mathcal{I}_{m^*+1}|\mathbb{P}(|B|<\tilde{C}_{m^*+1}\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{m^*}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{m^*}(2-q)}};\chi\right)} \\ &\quad +\frac{\sigma^{q-2}\phi\left(\tilde{C}_{m^*+1}\sqrt{\log(1/\sigma)}-\frac{1}{\log(1/\sigma)^{\mathcal{S}_0^{m^*+1}(2-q)}}\right)|\mathcal{I}_{m^*+1}|}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{m^*}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{m^*}(2-q)}};\chi\right)}. \end{aligned} \tag{3.26}$$

Again based on (3.25) It is clear that if $\tilde{C}_{m^*+1}$ is large enough, then the second term in (3.26) goes to zero. We now show the first term goes to zero as well:

$$\lim_{\sigma\to 0}\frac{\sigma^{q-2}\phi(0)|\mathcal{I}_{m^*+1}|\mathbb{P}(|B|<\tilde{C}_{m^*+1}\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{m^*}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{m^*}(2-q)}};\chi\right)}$$

$$\overset{(b)}{=} O(1)\cdot\lim_{\sigma\to 0}\frac{\sigma^{q-2}\frac{1}{\log(1/\sigma)^{\mathcal{S}_0^{m^*+1}(2-q)}}\sigma^{2-q+\epsilon}(\log(1/\sigma))^{\frac{2-q+\epsilon}{2}}}{\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{m^*+1}-\frac{\epsilon(2-q)}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_1^{m^*+1}(2-q)}}}$$

$$= O(1)\cdot\lim_{\sigma\to 0}\sigma^{\frac{\epsilon-(\epsilon+q-1)(2-q)^{m^*+1}}{q-1}}(\log(1/\sigma))^{\frac{-q+\epsilon}{2}} \overset{(c)}{=} 0,$$

where (b) holds from Lemma 3.4.6 and (c) is due to the condition we imposed on m^* that ensures $(2-q)^{m^*+1}\leq\frac{\epsilon}{\epsilon+q-1}$. The last remaining term of (3.21) is P_{m^*+2}. To prove $P_{m^*+2} \to 0$, we have

$$\begin{aligned} P_{m^*+2} &\leq \sigma^{q-2} \int_0^{\tilde{C}_{m^*+2}\sigma\sqrt{\log(1/\sigma)}} \int_{\mathcal{I}_{m^*+2}\backslash \mathcal{I}_{m^*+1}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}} \phi(z)dzdF(b) \\ &\quad + \sigma^{q-2} \int_{\tilde{C}_{m^*+2}\sigma\sqrt{\log(1/\sigma)}}^{\infty} \int_{\mathcal{I}_{m^*+2}\backslash \mathcal{I}_{m^*+1}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}} \phi(z)dzdF(b). \end{aligned}$$

By using the same strategy as we did for bounding P_i $(0 \leq i \leq m^*+1)$, the second integral above will go to zero as $\sigma \to 0$, when $\tilde{C}_{m^*+2}$ is chosen large enough. And the first integral can be bounded by

$$\frac{\sigma^{q-2}\phi(0)2\alpha\mathbb{P}(|B| \leq \tilde{C}_{m^*+2}\sigma\sqrt{\log(1/\sigma)})}{\eta_q^{2-q}\left(\frac{1}{\log(1/\sigma)^{\mathcal{S}_0^{m^*+1}(2-q)}};\chi\right)} \overset{(d)}{=} O(1)\sigma^{\epsilon}\log(1/\sigma)^{(2-q+\epsilon)/2+\mathcal{S}_1^{m^*+2}(2-q)} \to 0,$$

where (d) holds by Lemma 3.4.6 and the condition of Lemma 3.4.6 can be easily checked. This completes the proof. $\square$

Define

$$\mathcal{I}^{\gamma} \triangleq \left[-\frac{b}{\sigma} - \frac{\alpha}{\sigma^{1-\gamma}}, -\frac{b}{\sigma} + \frac{\alpha}{\sigma^{1-\gamma}}\right]. \tag{3.27}$$

In Lemma 3.4.7 we proved that:

$$\sigma^{q-2}\int_0^{\infty}\int_{\frac{-b}{\sigma}-\alpha}^{\frac{-b}{\sigma}+\alpha} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b) \to 0.$$

In the next lemma, we would like to extend this result and show that in fact,

$$\sigma^{q-2}\int_0^{\infty}\int_{\mathcal{I}^{\gamma}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b) \to 0.$$

Lemma 3.4.8. *For any given $q \in (1,2)$, suppose the conditions in Lemma 3.4.7 hold. Then for any fixed $0 < \gamma < 1$,*

$$\sigma^{q-2}\int_0^{\infty}\int_{\mathcal{I}^{\gamma}\backslash \mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b) \to 0,$$

as $\sigma \to 0$. Note that $\mathcal{I}_{m^*+2}$ is defined in (3.20).

Proof. As in the proof of Lemma 3.4.7, we break the integral into smaller subintervals and prove each one goes to zero. Consider the following intervals:

$$\mathcal{J}_i = \left[-\frac{b}{\sigma} - \frac{\alpha}{\sigma^{\frac{\epsilon}{1+\theta}\mathcal{S}_0^i(1-\epsilon)}}, -\frac{b}{\sigma} + \frac{\alpha}{\sigma^{\frac{\epsilon}{1+\theta}\mathcal{S}_0^i(1-\epsilon)}} \right],$$

where $\theta > 0$ is an arbitrarily small number and i is an arbitrary natural number. Note that $\{\mathcal{J}_i\}$ is a sequence of nested intervals and $\mathcal{I}_{m^*+2} \subset \mathcal{J}_0$. Our goal is to show that the following integrals go to zero as $\sigma \to 0$:

$$\begin{aligned} Q_{-1} &\triangleq \sigma^{q-2} \int_0^\infty \int_{\mathcal{J}_0 \backslash \mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} \phi(z) dz dF(b) \to 0, \\ Q_i &\triangleq \sigma^{q-2} \int_0^\infty \int_{\mathcal{J}_{i+1} \backslash \mathcal{J}_i} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} \phi(z) dz dF(b) \to 0, \quad i \geq 0. \end{aligned}$$

Define $\tilde{\sigma}_i \triangleq \frac{1}{\alpha}\sigma^{\frac{\epsilon}{1+\theta}\mathcal{S}_0^i(1-\epsilon)}$. Since $|b/\sigma + z| \geq \alpha$ for $z \notin \mathcal{I}_{m^*+2}$ we obtain

$$\begin{aligned} Q_{-1} &\leq \sigma^{q-2} \int_0^\infty \int_{\mathcal{J}_0 \backslash \mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(\alpha; \chi)|^{2-q}} \phi(z) dz dF(b) \\ &= \sigma^{q-2} \int_0^{\frac{\sigma}{\tilde{\sigma}_0}\log(1/\sigma)} \int_{\mathcal{J}_0 \backslash \mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(\alpha; \chi)|^{2-q}} \phi(z) dz dF(b) \\ &\quad + \sigma^{q-2} \int_{\frac{\sigma}{\tilde{\sigma}_0}\log(1/\sigma)}^\infty \int_{\mathcal{J}_0 \backslash \mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(\alpha; \chi)|^{2-q}} \phi(z) dz dF(b) \\ &\leq \sigma^{q-2} \int_0^{\frac{\sigma}{\tilde{\sigma}_0}\log(1/\sigma)} \frac{1}{|\eta_q(\alpha; \chi)|^{2-q}} dF(b) + \sigma^{q-2} \frac{\phi(\frac{\log(1/\sigma)}{\tilde{\sigma}_0} - \frac{1}{\tilde{\sigma}_0})|\mathcal{J}_0|}{|\eta_q(\alpha; \chi)|^{2-q}}. \end{aligned}$$

It is straightforward to notice that the second term above converges to zero. For the first term, by the condition $\mathbb{P}(|B| < t) = O(t^{2-q+\epsilon})$ we derive the following bounds

$$\begin{aligned} &\sigma^{q-2} \int_0^{\frac{\sigma}{\tilde{\sigma}_0}\log(1/\sigma)} \frac{1}{|\eta_q(\alpha; \chi)|^{2-q}} dF(b) \leq O(1)\sigma^{q-2}(\sigma\tilde{\sigma}_0^{-1}\log(1/\sigma))^{2-q+\epsilon} \\ &= O(1)\sigma^{\frac{\epsilon(q-1-\epsilon+\theta)}{1+\theta}} (\log(1/\sigma))^{2-q+\epsilon} \to 0. \end{aligned}$$

Now we discuss the term Q_i for $i \geq 0$. Similarly as we bounded Q_{-1} we have

$$\begin{aligned} Q_i &\leq \sigma^{q-2} \int_0^{\frac{\sigma}{\tilde{\sigma}_{i+1}}\log(1/\sigma)} \int_{\mathcal{J}_{i+1} \backslash \mathcal{J}_i} \frac{1}{|\eta_q(\frac{1}{\tilde{\sigma}_i}; \chi)|^{2-q}} \phi(z) dz dF(b) \\ &\quad + \sigma^{q-2} \int_{\frac{\sigma}{\tilde{\sigma}_{i+1}}\log(1/\sigma)}^\infty \int_{\mathcal{J}_{i+1} \backslash \mathcal{J}_i} \frac{1}{|\eta_q(\frac{1}{\tilde{\sigma}_i}; \chi)|^{2-q}} \phi(z) dz dF(b). \end{aligned} \tag{3.28}$$

The second integral in (3.28) can be easily shown convergent to zero as $\sigma \to 0$. We now focus on the first integral.

$$\begin{aligned}
&\sigma^{q-2}\int_0^{\frac{\sigma}{\tilde{\sigma}_{i+1}}\log(1/\sigma)}\int_{\mathcal{J}_{i+1}\backslash\mathcal{J}_i}\frac{1}{|\eta_q(\frac{1}{\tilde{\sigma}_i};\chi)|^{2-q}}\phi(z)dzdF(b)\\
\leq\ & \frac{\sigma^{q-2}}{|\eta_q(\frac{1}{\tilde{\sigma}_i};\chi)|^{2-q}}\mathbb{P}\big(|B|\leq\frac{\sigma}{\tilde{\sigma}_{i+1}}\log(1/\sigma)\big)\\
\leq\ & O(1)\frac{\sigma^{\epsilon}(\log(1/\sigma))^{2-q+\epsilon}}{|\eta_q(\frac{1}{\tilde{\sigma}_i};\chi)|^{2-q}\tilde{\sigma}_{i+1}^{2-q+\epsilon}}\overset{(a)}{=}O(1)\frac{\sigma^{\epsilon}(\log(1/\sigma))^{2-q+\epsilon}\tilde{\sigma}_i^{2-q}}{\tilde{\sigma}_{i+1}^{2-q+\epsilon}}\\
=\ & O(1)\sigma^{\frac{\epsilon(\theta+(q-1-\epsilon)(1-\epsilon)^{i+1})}{1+\theta}}(\log(1/\sigma))^{2-q+\epsilon}=o(1).
\end{aligned}$$

We have used Lemma 3.4.6 to obtain (a). Above all we have showed that for any given natural number $i\geq 0$,

$$\lim_{\sigma\to 0}\sigma^{q-2}\int_0^{\infty}\int_{\mathcal{J}_i\backslash\mathcal{I}_{m^*+2}}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)=0.$$

Now note that as i goes to infinity, the exponent of σ in the interval $\mathcal{J}_i$ goes to $\frac{\epsilon}{1+\theta}(1+(1-\epsilon)+(1-\epsilon)^2+\ldots)=\frac{1}{1+\theta}$. So, by choosing small enough θ and sufficiently large i we can make $\mathcal{I}^{\gamma}\subset\mathcal{J}_i$, hence completing the proof. □

In the last two lemmas, we have been able to prove that for $\chi=C\sigma^q$,

$$\sigma^{q-2}\int_0^{\infty}\int_{\mathcal{I}^{\gamma}}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)\to 0.$$

This result will be used to characterize the following limit

$$\lim_{\sigma\to 0}\sigma^{q-2}\int_0^{\infty}\int_{\mathbb{R}}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b).$$

Before that we mention a simple lemma that will be applied several times in our proofs.

Lemma 3.4.9. *For $1<q<2$ we have*

$$\frac{1}{|\eta_q(u;\chi)|^{2-q}+\chi q(q-1)}\leq\frac{2}{|u|^{2-q}(q-1)}.$$

Proof. It is sufficient to consider $u>0$. We analyze two different cases:

1. $\chi \leq u^{2-q}\frac{1}{2q}$: According to Lemma 3.4.1 part (ii), since we know $\eta_q(u;\chi) \leq u$, we have

$$\eta_q(u;\chi) = u - \chi q \eta_q^{q-1}(u;\chi) \geq u - \chi q u^{q-1} \geq u - u^{2-q}\frac{1}{2q} q \alpha^{q-1} = \frac{u}{2}.$$

Hence,

$$\frac{1}{|\eta_q(u;\chi)|^{2-q} + \chi q(q-1)} \leq \frac{1}{|\eta_q(u;\chi)|^{2-q}} \leq \frac{2^{2-q}}{u^{2-q}} \leq \frac{2}{(q-1)u^{2-q}}.$$

2. $\chi \geq u^{2-q}\frac{1}{2q}$:

$$\frac{1}{|\eta_q(u;\chi)|^{2-q} + \chi q(q-1)} \leq \frac{1}{\chi q(q-1)} \leq \frac{2}{(q-1)u^{2-q}}.$$

This completes our proof. □

Now we can consider one of the main results of this section.

Lemma 3.4.10. *For any given $q \in (1,2)$, suppose the conditions in Lemma 3.4.7 hold. Then for $\chi = C\sigma^q$ we have*

$$\lim_{\sigma \to 0} \frac{R_q(\chi,\sigma) - 1}{\sigma^2} = C^2 q^2 \mathbb{E}|B|^{2q-2} - 2Cq(q-1)\mathbb{E}|B|^{q-2}.$$

Proof. We follow the same roadmap as in the proof of Lemma 3.4.3. Recall that

$$R_q(\chi,\sigma) - 1 = \underbrace{\chi^2 q^2 \mathbb{E}|\eta_q(B/\sigma + Z;\chi)|^{2q-2}}_{S_1} \underbrace{-2\chi q(q-1)\mathbb{E}\frac{|\eta_q(B/\sigma + Z;\chi)|^{q-2}}{1 + \chi q(q-1)|\eta_q(B/\sigma + Z;\chi)|^{q-2}}}_{S_2}. \tag{3.29}$$

The first term S_1 can be calculated in the same way as in the proof of Lemma 3.4.3.

$$\lim_{\sigma \to 0} \sigma^{-2} S_1 = C^2 q^2 \mathbb{E}|B|^{2q-2}. \tag{3.30}$$

We now focus on analyzing S_2. First note that restricting $|B|$ to be bounded away from 0 makes it possible to follow the same arguments used in the proof of Lemma 3.4.3 to obtain,

$$\lim_{\sigma \to 0} \mathbb{E}\frac{\mathbb{1}(|B| > 1)}{|\eta_q(|B| + \sigma Z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)} = \mathbb{E}|B|^{q-2}\mathbb{1}(|B| > 1). \tag{3.31}$$

Hence we next consider the event $|B| \leq 1$.

$$
\begin{aligned}
&\mathbb{E}\frac{\mathbb{1}(|B| \leq 1)}{|\eta_q(|B| + \sigma Z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)} \\
= &\underbrace{\int_0^1 \int_{-b/\sigma - b^c/(2\sigma)}^{-b/\sigma + b^c/(2\sigma)} \frac{1}{|\eta_q(b + \sigma z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)} \phi(z) dz dF(b)}_{T_1} \\
&+ \underbrace{\int_0^1 \int_{\mathbb{R}\setminus[-b/\sigma - b^c/(2\sigma), -b/\sigma + b^c/(2\sigma)]} \frac{1}{|\eta_q(b + \sigma z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)} \phi(z) dz dF(b)}_{T_2},
\end{aligned}
$$

where $c > 1$ is a constant that we will specify later. We first analyze T_2. Note that,

$$
T_2 = \mathbb{E}\frac{\mathbb{1}(|B + \sigma Z| \geq |B|^c/2, |B| \leq 1)}{|\eta_q(B + \sigma Z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)},
$$

and

$$
\frac{\mathbb{1}(|B + \sigma Z| \geq |B|^c/2, |B| \leq 1)}{|\eta_q(B + \sigma Z; C\sigma^2)|^{2-q} + C\sigma^2 q(q-1)} \overset{(a)}{\leq} \frac{2\mathbb{1}(|B + \sigma Z| \geq |B|^c/2)}{(q-1)|B + \sigma Z|^{2-q}} \leq \frac{|B|^{c(q-2)}}{2^{q-3}(q-1)},
$$

where (a) is due to Lemma 3.4.9. For any $1 < q < 2$, it is straightforward to verify that $\mathbb{E}|B|^{c(q-2)} < \infty$ if c is chosen close enough to 1. We can then apply Dominated Convergence Theorem (DCT) to obtain

$$
\lim_{\sigma \to 0} T_2 = \mathbb{E}\mathbb{1}(|B| \geq |B|^c/2, |B| \leq 1)|B|^{q-2} = \mathbb{E}|B|^{q-2}\mathbb{1}(|B| \leq 1). \tag{3.32}
$$

We now turn to bounding T_1. According to Lemmas 3.4.7 and 3.4.8, we know

$$
\sigma^{q-2} \int_0^\infty \int_{\mathcal{I}^\gamma} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} \phi(z) dz dF(b) \to 0,
$$

where $\mathcal{I}^\gamma = [-\frac{b}{\sigma} - \frac{\alpha}{\sigma^{1-\gamma}}, -\frac{b}{\sigma} + \frac{\alpha}{\sigma^{1-\gamma}}]$. Define $\mathcal{I}_c^\gamma = [-\frac{b}{\sigma} - \frac{b^c}{\sigma^{1-\gamma}}, -\frac{b}{\sigma} + \frac{b^c}{\sigma^{1-\gamma}}]$ and $\tilde{\mathcal{I}}^c = [-\frac{b}{\sigma} - \frac{b^c}{2\sigma}, -\frac{b}{\sigma} + \frac{b^c}{2\sigma}]$. For $0 \leq b \leq 1$, we get $\mathcal{I}_c^\gamma \subseteq \mathcal{I}^\gamma$ for any given $\alpha > 1$. Therefore,

$$
T_3 \triangleq \sigma^{q-2} \int_0^1 \int_{\mathcal{I}_c^\gamma} \frac{1}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} \phi(z) dz dF(b) \to 0.
$$

Hence to bound T_1, it is sufficient to bound $T_1 - T_3$:

$$\begin{aligned}
T_1 - T_3 &= \sigma^{q-2}\int_0^1\int_{\tilde{\mathcal{I}}^c\backslash\mathcal{I}_c^\gamma}\frac{1}{|\eta_q(b/\sigma+z;\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)\\
&\leq \sigma^{q-2}\int_0^1\int_{\tilde{\mathcal{I}}^c\backslash\mathcal{I}_c^\gamma}\frac{1}{|\eta_q(b^c/\sigma^{1-\gamma};\chi)|^{2-q}+\chi q(q-1)}\phi(z)dzdF(b)\\
&\overset{(b)}{\leq} \sigma^{q-2+(1-\gamma)(2-q)}\int_0^1\int_{\tilde{\mathcal{I}}^c\backslash\mathcal{I}_c^\gamma}\frac{2b^{c(q-2)}}{q-1}\phi(z)dzdF(b)\\
&= \underbrace{\sigma^{q-2+(1-\gamma)(2-q)}\int_0^{\tilde{C}\sigma\sqrt{\log(1/\sigma)}}\int_{\tilde{\mathcal{I}}^c\backslash\mathcal{I}_c^\gamma}\frac{2b^{c(q-2)}}{q-1}\phi(z)dzdF(b)}_{T_4}\\
&\quad \underbrace{+\,\sigma^{q-2+(1-\gamma)(2-q)}\int_{\tilde{C}\sigma\sqrt{\log(1/\sigma)}}^1\int_{\tilde{\mathcal{I}}^c\backslash\mathcal{I}_c^\gamma}\frac{2b^{c(q-2)}}{q-1}\phi(z)dzdF(b)}_{T_5},
\end{aligned}$$

where (b) is the result of Lemma 3.4.9 and $\tilde{C}$ is a positive constant. We first bound T_5 in the following:

$$T_5 \leq \frac{2\sigma^{q-2+(1-\gamma)(2-q)}}{q-1}\int_{\tilde{C}\sigma\sqrt{\log(1/\sigma)}}^1\frac{b^{c(q-1)}}{\sigma}\phi\left(\frac{b}{2\sigma}\right)dF(b) \leq \frac{2\sigma^{q-3+(1-\gamma)(2-q)}}{q-1}\phi(\tilde{C}\sqrt{\log(1/\sigma)}/2).$$

It is then easily seen that T_5 goes to zero by choosing large enough $\tilde{C}$. For the remaining term T_4,

$$\begin{aligned}
T_4 &\leq \frac{2\sigma^{q-3+(1-\gamma)(2-q)}}{q-1}\int_0^{\tilde{C}\sigma\sqrt{\log(1/\sigma)}}b^{c(q-1)}\phi\left(\frac{b}{2\sigma}\right)dF(b)\\
&\leq \frac{2\sigma^{q-3+(1-\gamma)(2-q)}}{q-1}(\tilde{C}\sigma\sqrt{\log(1/\sigma)})^{c(q-1)}\phi(0)\mathbb{P}(|B|\leq\tilde{C}\sigma\sqrt{\log(1/\sigma)})\\
&\leq O(1)\sigma^{c(q-1)-\gamma(2-q)+1-q+\epsilon}(\log(1/\sigma))^{(c(q-1)+2-q+\epsilon)/2}\to 0.
\end{aligned}$$

To obtain the last statement, we can choose γ close enough to zero and c close to 1. Hence we can conclude $T_1 \to 0$ as $\sigma \to 0$. This combined with the results in (3.31) and (3.32) gives us

$$\lim_{\sigma\to 0} -\sigma^{-2}S_2 = 2Cq(q-1)\mathbb{E}\frac{1}{|\eta_q(|B|+\sigma Z;C\sigma^2)|^{2-q}+C\sigma^2q(q-1)} = 2Cq(q-1)\mathbb{E}|B|^{q-2}.$$

The above result together with (3.30) finishes the proof. □

As stated in the roadmap of the proof, our first goal is to characterize the convergence rate of $\chi_q^*(\sigma)$. Towards this goal, we first show that $\chi_q^*(\sigma)$ cannot be either too large or too small. In particular, in Lemmas 3.4.11 and 3.4.12, we show that $\chi_q^*(\sigma) = O(\sigma^{q-1})$ and $\chi_q^*(\sigma) = \Omega(\sigma^q)$. We then utilize such result in Lemma 3.4.14 to conclude that $\chi_q^*(\sigma) = \Theta(\sigma^q)$.

Lemma 3.4.11. *Suppose $\mathbb{E}|B|^2 < \infty$, if $\chi\sigma^{1-q} = \infty$ and $\chi = o(1)$, then $R_q(\chi, \sigma) \to \infty$, as $\sigma \to 0$.*

Proof. Consider the formula of $R_q(\chi, \sigma)$ in (3.29). Since $\chi = o(1)$, it is straightforward to apply Dominated Convergence Theorem to obtain

$$\lim_{\sigma \to 0} \chi^{-2}\sigma^{2q-2} S_1 = q^2 \mathbb{E}|B|^{2q-2}.$$

Because $\chi^2\sigma^{2-2q} \to \infty$, we know $S_1 \to \infty$. Also note

$$|S_2| \leq 2\chi q(q-1) \cdot \frac{1}{\chi q(q-1)} = 2.$$

Hence, $R_q(\chi, \sigma) \to \infty$. □

Lemma 3.4.12. *Suppose that the same conditions for B in Lemma 3.4.10 hold, if $\chi = o(\sigma^q)$, then*

$$\frac{R_q(\chi, \sigma) - 1}{\sigma^2} \to 0, \quad \textit{as } \sigma \to 0.$$

Proof. Consider the expression of $R_q(\chi, \sigma) - 1$ in (3.29). First note that

$$\lim_{\sigma \to 0} \frac{S_1}{\sigma^2} = \lim_{\sigma \to 0} \frac{\chi^2 \sigma^{2-2q} q^2 \mathbb{E}|\eta_q(B + \sigma Z; \chi\sigma^{2-q})|^{2q-2}}{\sigma^2} = 0.$$

Now we study the behavior of S_2. Recall that we defined $\mathcal{I}_{-1} = \left[-\frac{b}{\sigma} - \frac{\sigma^{q-\epsilon}}{\log(\frac{1}{\sigma})}, -\frac{b}{\sigma} + \frac{\sigma^{q-\epsilon}}{\log(\frac{1}{\sigma})}\right]$ and $\mathcal{I}^\gamma = \left[-\frac{b}{\sigma} - \frac{\alpha}{\sigma^{1-\gamma}}, -\frac{b}{\sigma} + \frac{\alpha}{\sigma^{1-\gamma}}\right]$ in (3.20) and (3.27), respectively. It is straightforward to use the same argument as for bounding P_{-1} in the proof of Lemma 3.4.7 (see the derivations in (3.22)) to have

$$\frac{\chi}{\sigma^2}\int_0^\infty \int_{\mathcal{I}_{-1}} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) \leq \frac{\chi}{\sigma^2}\int_0^\infty \int_{\mathcal{I}_{-1}} \frac{\phi(z)}{\chi q(q-1)} dz dF(b) \to 0.$$

Moreover, since $\chi < C\sigma^q$ for small enough σ, Lemma 3.4.1 part (v) implies

$$|\eta_q(b/\sigma + z; \chi)| \geq |\eta_q(b/\sigma + z; C\sigma^q)|.$$

Therefore, as $\sigma \to 0$

$$\begin{aligned} &\frac{\chi}{\sigma^2} \int_0^{\infty} \int_{\mathcal{I}^{\gamma} \backslash \mathcal{I}_{-1}} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q}} dz dF(b) \\ &\leq \frac{\chi}{\sigma^q} \cdot \frac{1}{\sigma^{2-q}} \int_0^{\infty} \int_{\mathcal{I}^{\gamma} \backslash \mathcal{I}_{-1}} \frac{\phi(z)}{|\eta_q(b/\sigma + z; C\sigma^q)|^{2-q}} dz dF(b) \to 0, \end{aligned}$$

where the last statement holds because of $\frac{1}{\sigma^{2-q}} \int_0^{\infty} \int_{\mathcal{I}^{\gamma} \backslash \mathcal{I}_{-1}} \frac{\phi(z)}{|\eta_q(b/\sigma + z; C\sigma^q)|^{2-q}} dz dF(b) \to 0$ that has already been shown in the proof of Lemmas 3.4.7 and 3.4.8. Above all we have proved

$$\frac{\chi}{\sigma^2} \int_0^{\infty} \int_{\mathcal{I}^{\gamma}} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) \to 0.$$

Based on the above result, we can easily follow the same derivations of bounding the term T_1 in the proof of Lemma 3.4.10 to conclude

$$\lim_{\sigma \to 0} \frac{\chi}{\sigma^2} \int_0^1 \int_{-b/\sigma - b^c/(2\sigma)}^{-b/\sigma + b^c/(2\sigma)} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) = 0. \tag{3.33}$$

Furthermore, because $\chi = o(\sigma^q)$, the analyses to derive Equation (3.31) and bound T_2 in the proof of Lemma 3.4.10 can be adapted here and yield

$$\begin{aligned} &\lim_{\sigma \to 0} \frac{\chi}{\sigma^2} \int_1^{\infty} \int_{-\infty}^{+\infty} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) = 0, \qquad (3.34) \\ &\lim_{\sigma \to 0} \frac{\chi}{\sigma^2} \int_0^1 \int_{\mathbb{R} \backslash [-b/\sigma - b^c/(2\sigma), -b/\sigma + b^c/(2\sigma)]} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) = 0. \end{aligned}$$

Putting results (3.33) and (3.34) together gives us

$$\lim_{\sigma \to 0} \frac{-S_2}{\sigma^2} = \lim_{\sigma \to 0} \frac{2\chi q(q-1)}{\sigma^2} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{\phi(z)}{|\eta_q(b/\sigma + z; \chi)|^{2-q} + \chi q(q-1)} dz dF(b) = 0.$$

This finishes the proof. □

Collecting the results from Lemmas 3.4.10, 3.4.11 and 3.4.12, we can upper and lower bound the optimal threshold value $\chi_q^*(\sigma)$ as shown in the following corollary.

Corollary 3.4.13. *Suppose the conditions for B in Lemma 3.4.10 hold. Then as $\sigma \to 0$, we have*

$$\chi_q^*(\sigma) = \Omega(\sigma^q), \quad \chi_q^*(\sigma) = O(\sigma^{q-1}).$$

Proof. Since $\chi = \chi_q^*(\sigma)$ minimizes $R_q(\chi, \sigma)$, we know

$$R_q(\chi_q^*(\sigma), \sigma) \le R_q(0, \sigma) = 1, \quad \text{for any } \sigma > 0, \tag{3.35}$$

$$\sigma^{-1}(R_q(\chi_q^*(\sigma), \sigma) - 1) \le \sigma^{-2}(R_q(C\sigma^q, \sigma) - 1) < -c, \quad \text{for small enough } \sigma, \tag{3.36}$$

where the last inequality is due to Lemma 3.4.10 with an appropriate choice of C, and c is a positive constant. Note that we already know $\chi_q^*(\sigma) = o(1)$. If $\chi_q^*(\sigma) \neq O(\sigma^{q-1})$, Lemma 3.4.11 will contradict with (3.35). If $\chi_q^*(\sigma) \neq \Omega(\sigma^q)$, Lemma 3.4.12 will contradict with (3.36). □

We are now able to derive the exact convergence rate of $\chi_q^*(\sigma)$ and $R_q(\chi_q^*(\sigma), \sigma)$.

Lemma 3.4.14. *For any given $q \in (1, 2)$, suppose the conditions in Lemma 3.4.7 for B hold. Then we have*

$$\begin{aligned} \chi_q^*(\sigma) &= \frac{(q-1)\mathbb{E}|B|^{q-2}}{q\mathbb{E}|B|^{2q-2}}\sigma^q + o(\sigma^q), \\ R_q(\chi_q^*(\sigma), \sigma) &= 1 - \frac{(q-1)^2(\mathbb{E}|B|^{q-2})^2}{\mathbb{E}|B|^{2q-2}}\sigma^2 + o(\sigma^2). \end{aligned}$$

Proof. In this proof, we use χ^* to denote $\chi_q^*(\sigma)$ for notational simplicity. Using the notations in Equation (3.13), we know that χ^* satisfies the following equation:

$$0 = \chi^* U_1 - U_2 - \chi^* U_3.$$

Our first goal is to show that $\sigma^{q-2}U_2 \to q(q-1)\mathbb{E}|B|^{q-2}$ as $\sigma \to 0$. Define the interval

$$\mathcal{K} = [-b/\sigma - (\chi^*)^{1/(2-q)}, -b/\sigma + (\chi^*)^{1/(2-q)}]. \tag{3.37}$$

Then we have,

$$\begin{aligned} \frac{U_2}{\sigma^{2-q}} &= \frac{q(q-1)}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \\ &+ \frac{q(q-1)}{\sigma^{2-q}} \int_0^\infty \int_{\mathbb{R}\backslash\mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b). \end{aligned} \tag{3.38}$$

We first show that the first term in (3.38) goes to zero. Note that $\eta_q((\chi^*)^{1/(2-q)};\chi^*) = (\chi^*)^{1/(2-q)}\eta_q(1;1)$ by Lemma 3.4.1 part (iii), we thus have

$$\begin{aligned}
&\frac{q(q-1)}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{|\eta_q(b/\sigma+z;\chi^*)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi^*)|^{2-q}+\chi^*q(q-1))^3}\phi(z)dzdF(b)\\
\leq\ &\frac{q(q-1)}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{|\eta_q((\chi^*)^{1/(2-q)};\chi^*)|^{4-2q}}{(\chi^*q(q-1))^3}\phi(z)dzdF(b)\\
\leq\ &\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{\eta_q^{4-2q}(1;1)}{\chi^*(q(q-1))^2}\phi(z)dzdF(b)\\
\leq\ &\frac{1}{\sigma^{2-q}}\int_0^{C_1\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{K}}\frac{\eta_q^{4-2q}(1;1)}{\chi^*(q(q-1))^2}\phi(z)dzdF(b)\\
&+\frac{\eta_q^{4-2q}(1;1)}{q^2(q-1)^2\sigma^{2-q}\chi^*}|\mathcal{K}|\phi(C_1\sqrt{\log(1/\sigma)}-(\chi^*)^{1/(2-q)}).
\end{aligned}$$

Since we have already shown $\chi^*=\Omega(\sigma^q)$ in Corollary 3.4.13, it is straightforward to see that the second integral in the above bound is negligible for large enough C_1. For the first term, we know

$$\frac{1}{\chi^*\sigma^{2-q}}\int_0^{C_1\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{K}}\phi(z)dzdF(b)\leq O(1)(\chi^*)^{(q-1)/(2-q)}\sigma^{\epsilon}(\log(1/\sigma))^{\frac{2-q+\epsilon}{2}}=o(1).$$

Our next goal is to find the limit of the second term in (3.38). In order to do that, we again break the integral into several pieces. Recall the intervals $\mathcal{I}^\gamma,\mathcal{I}_{-1},\mathcal{I}_0,\mathcal{I}_1,\ldots,\mathcal{J}_0,\mathcal{J}_1,\ldots$ that we introduced in Lemmas 3.4.7 and 3.4.8. We consider two different cases:

1. In this case, we assume that $(\chi^*)^{1/(2-q)}=o(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})$.

   Hence $\mathcal{K}\subseteq\mathcal{I}_{-1}$. We have

$$\begin{aligned}
&\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{I}_{-1}\backslash\mathcal{K}}\frac{|\eta_q(b/\sigma+z;\chi^*)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi^*)|^{2-q}+\chi^*q(q-1))^3}\phi(z)dzdF(b)\\
\leq\ &\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{I}_{-1}\backslash\mathcal{K}}\frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}}\phi(z)dzdF(b)\\
\leq\ &\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{I}_{-1}\backslash\mathcal{K}}\frac{1}{\chi^*\eta_q^{2-q}(1;1)}\phi(z)dzdF(b)\\
\leq\ &\frac{1}{\sigma^{2-q}}\int_0^{C_2\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{I}_{-1}\backslash\mathcal{K}}\frac{1}{\chi^*\eta_q^{2-q}(1;1)}\phi(z)dzdF(b)\\
&+\frac{1}{\sigma^{2-q}}\int_{C_2\sigma\sqrt{\log(1/\sigma)}}^\infty\int_{\mathcal{I}_{-1}\backslash\mathcal{K}}\frac{1}{\chi^*\eta_q^{2-q}(1;1)}\phi(z)dzdF(b).
\end{aligned}$$

The fact that $\chi^*(\sigma) = \Omega(\sigma^q)$ enables us to conclude that the second integral above goes to zero by choosing large enough C_2. Regarding the first term we know

$$
\begin{aligned}
&\frac{1}{\chi^* \sigma^{2-q}} \int_0^{C_2\sigma\sqrt{\log(1/\sigma)}} \int_{\mathcal{I}_{-1}\backslash\mathcal{K}} \phi(z) dz dF(b) \\
\leq\ & \frac{\phi(0)|\mathcal{I}_{-1}|\mathbb{P}(|B| \leq C_2\sigma\sqrt{\log(1/\sigma)})}{\sigma^{2-q}\chi^*} = O(1)\cdot \frac{\sigma^q}{\chi^*}\cdot (\log(1/\sigma))^{\frac{-q+\epsilon}{2}} \overset{(a)}{=} o(1),
\end{aligned}
$$

where (a) is due to $\chi^* = \Omega(\sigma^q)$. We now consider another integral.

$$
\begin{aligned}
&\frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma\backslash\mathcal{I}_{-1}} \frac{|\eta_q(b/\sigma+z;\chi^*)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \\
\leq\ & \frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma\backslash\mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b).
\end{aligned}
$$

Our goal is to show that this integral goes to zero as well. We use the following calculations:

$$
\begin{aligned}
&\frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma\backslash\mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b) \\
\leq\ & \frac{1}{\sigma^{2-q}} \sum_{i=0}^{m_*+2} \int_0^\infty \int_{\mathcal{I}_i\backslash\mathcal{I}_{i-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b) \\
&+ \frac{1}{\sigma^{2-q}} \sum_{i=1}^{\ell} \int_0^\infty \int_{\mathcal{J}_i\backslash\mathcal{J}_{i-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b) \\
&+ \frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{J}_0\backslash\mathcal{I}_{m^*+2}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b),
\end{aligned}
$$

where ℓ is chosen in a way such that $\mathcal{I}^\gamma \subseteq \mathcal{J}_\ell$. Define $m_i = |\mathcal{I}_i|$ and $\tilde{m}_i = |\mathcal{J}_i|$. Note that we did similar calculations for the case $\chi = C\sigma^q$ in Lemmas 3.4.7 and 3.4.8. The key argument regarding χ that we used there to show each term above converges to zero was that $\eta_q(m_i; C\sigma^q) = \Theta(m_i)$ and $\eta_q(\tilde{m}_i; C\sigma^q) = \Theta(\tilde{m}_i)$. Hence, if we can show that $\eta_q(m_i;\chi^*) = \Theta(m_i)$ and $\eta_q(\tilde{m}_i;\chi^*) = \Theta(\tilde{m}_i)$ in the current case, then those proofs will carry over and we will have

$$
\frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma\backslash\mathcal{I}_{-1}} \frac{1}{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}} \phi(z) dz dF(b) \to 0.
$$

For this purpose, we make use of Lemma 3.4.6. Note that since $m_{-1} < m_0 < m_1 < \ldots < m_{m*+2} < \tilde{m}_0 < \tilde{m}_1 < \tilde{m}_2 \ldots < \tilde{m}_\ell$, we only need to confirm the condition of Lemma 3.4.6 for m_{-1}. We have

$$\chi^* m_{-1}^{q-2} = \chi^* \sigma^{(q-\epsilon)(q-2)} \log(1/\sigma)^{2-q} = o(1),$$

by the assumption of Case 1. Hence in the current case we have obtained

$$\frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \to 0.$$

Furthermore, it is clear that

$$\frac{\sigma^{q-2}|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \leq \frac{1}{|\eta_q(b + \sigma z; \chi^* \sigma^{2-q})|^{2-q} + \chi^* \sigma^{2-q} q(q-1)}.$$

We can then follow the same line of arguments for deriving $\lim_{\sigma \to 0} -S_2/\sigma^2$ in the proof of Lemma 3.4.10 to obtain $\lim_{\sigma \to 0} \sigma^{q-2} U_2 = q(q-1)\mathbb{E}|B|^{q-2}$.

2. The other case is $(\chi^*)^{\frac{1}{2-q}} = \Omega(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})$. Because $(\chi^*)^{\frac{1}{2-q}} = \Omega(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})$ and $\chi^* = O(\sigma^{q-1})$, there exists a value of $0 \leq \bar{m} \leq m^* + 1$ such that for σ small enough, $(\chi^*)^{\frac{1}{2-q}} = o(|\mathcal{I}_{\bar{m}}|)$ and $(\chi^*)^{\frac{1}{2-q}} = \Omega(|\mathcal{I}_{\bar{m}-1}|)$. We then break the integral into:

$$\begin{aligned}
&\frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma \backslash \mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \\
&= \frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}_{\bar{m}} \backslash \mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \\
&+ \frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma \backslash \mathcal{I}_{\bar{m}}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \quad (3.39)
\end{aligned}$$

Once we show that each of the two integrals above goes to zero as $\sigma \to 0$, then the subsequent arguments will be exactly the same as the ones in Case 1.

Regarding the first integral,

$$\begin{aligned}
&\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{I}_{\bar m}\backslash\mathcal{K}}\frac{|\eta_q(b/\sigma+z;\chi^*)|^{4-2q}}{(|\eta_q(b/\sigma+z;\chi^*)|^{2-q}+\chi^*q(q-1))^3}\phi(z)dzdF(b)\\
\le\;& \frac{1}{\sigma^{2-q}}\int_0^{C_3\sigma\sqrt{\log(1/\sigma)}}\int_{\mathcal{I}_{\bar m}\backslash\mathcal{K}}\frac{1}{\chi^*\eta_q^{2-q}(1;1)}\phi(z)dzdF(b)\\
&+\frac{1}{\sigma^{2-q}}\int_{C_3\sigma\sqrt{\log(1/\sigma)}}^\infty\int_{\mathcal{I}_{\bar m}\backslash\mathcal{K}}\frac{1}{\chi^*\eta_q^{2-q}(1;1)}\phi(z)dzdF(b)\\
\le\;& \frac{\phi(0)|\mathcal{I}_{\bar m}|\mathbb{P}(|B|\le C_3\sigma\sqrt{\log(1/\sigma)})}{\sigma^{2-q}\eta_q^{2-q}(1;1)\chi^*}+\frac{\phi(C_3\sqrt{\log(1/\sigma)}/2)}{\sigma^{2-q}\eta_q^{2-q}(1;1)\chi^*}.
\end{aligned}$$

Since $\chi^*=\Omega(\sigma^q)$ from Corollary 3.4.13, it is clear that the second term in the above upper bound goes to zero by choosing large enough C_3. Regarding the first term we have

$$\begin{aligned}
\frac{|\mathcal{I}_{\bar m}|\mathbb{P}(|B|\le C_3\sigma\sqrt{\log(1/\sigma)})}{\sigma^{2-q}\chi^*} &\overset{(a)}{\le} O(1)\frac{\sigma^\epsilon(\log(1/\sigma))^{\frac{2-q+\epsilon}{2}}\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{\bar m}-\frac{\epsilon}{q-1}}}{\chi^*\log(1/\sigma)^{\mathcal{S}_0^{\bar m}(2-q)}}\\
&\overset{(b)}{\le}\begin{cases} O(1)(\log(1/\sigma))^{\frac{-q+\epsilon}{2}}=o(1) & \text{if } \bar m>0,\\ O(1)(\log(1/\sigma))^{\frac{\epsilon+4-3q}{2}}\sigma^{q^2-(2+\epsilon)q+3\epsilon+1}=o(1) & \text{if } \bar m=0,\end{cases}
\end{aligned}$$

where (a) holds even when $\bar m=m^*+1$ since $\frac{\epsilon+q-1}{q-1}(2-q)^{\bar m}-\frac{\epsilon}{q-1}\le 0$ by the definition of m^*; and (b) is due to the fact that $(\chi^*)^{1/(2-q)}=\Omega\left(\frac{\sigma^{\frac{\epsilon+q-1}{q-1}(2-q)^{\bar m-1}-\frac{\epsilon}{q-1}}}{(\log(1/\sigma))^{\mathcal{S}_0^{\bar m-1}(2-q)}}\right)$ when $\bar m>0$ and $(\chi^*)^{1/(2-q)}=\Omega(\frac{\sigma^{q-\epsilon}}{\log(1/\sigma)})$ when $\bar m=0$, according to the choice of $\bar m$. For the second integral in (3.39), note that $(\chi^*)^{\frac{1}{2-q}}=o(|\mathcal{I}_{\bar m}|)$, hence $\chi^*|\mathcal{I}_{\bar m}|^{q-2}\to 0$. It implies that the arguments in calculating the second integral in Case 1 hold here as well.

So far we have been able to derive the limit of $\sigma^{q-2}U_2$. We next analyze the term $\sigma^{q-2}\chi^*U_3$ and show that it goes to zero as $\sigma\to 0$. We have

$$\begin{aligned}
&\frac{\chi^*}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{|\eta_q(b/\sigma+z;\chi^*)|^{2-q}}{(|\eta_q(b/\sigma+z;\chi^*)|^{2-q}+\chi^*q(q-1))^3}\phi(z)dzdF(b)\le\\
&\frac{\chi^*}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{\chi^*\eta_q^{2-q}(1;1)}{(\chi^*q(q-1))^3}\phi(z)dzdF(b)=\frac{1}{\sigma^{2-q}}\int_0^\infty\int_{\mathcal{K}}\frac{\eta_q^{2-q}(1;1)}{\chi^*(q(q-1))^3}\phi(z)dzdF(b)
\end{aligned}$$

The upper bound above has been shown to be zero in the preceding calculations regarding the first term in (3.38). Furthermore, note that when $z \notin \mathcal{K}$,

$$|\eta_q(b/\sigma + z; \chi^*)|^{2-q} \geq \chi^* \eta_q^{2-q}(1;1).$$

We can then obtain

$$\begin{aligned} &\frac{\chi^*}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma \setminus \mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{2-q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) \\ &\quad \leq \ O(1) \cdot \frac{1}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma \setminus \mathcal{K}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{4-2q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b). \end{aligned}$$

The last term has been shown to converge to zero in the analysis of $\sigma^{q-2} U_2$. Above all we have derived that

$$\lim_{\sigma \to 0} \frac{\chi^*}{\sigma^{2-q}} \int_0^\infty \int_{\mathcal{I}^\gamma} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{2-q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) = 0.$$

This together with the fact

$$\frac{\chi^* |\eta_q(b/\sigma + z; \chi^*)|^{2-q}}{\sigma^{2-q}(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \leq \frac{q^{-1}(q-1)^{-1}}{|\eta_q(b + \sigma z; \sigma^{2-q}\chi^*)|^{2-q} + \sigma^{2-q}\chi^* q(q-1)},$$

we can again follow the line of arguments for $-\sigma^{-2} S_2$ in the proof of Lemma 3.4.10 to get

$$\lim_{\sigma \to 0} \frac{\chi^*}{\sigma^{2-q}} \int_0^\infty \int_{\mathbb{R}} \frac{|\eta_q(b/\sigma + z; \chi^*)|^{2-q}}{(|\eta_q(b/\sigma + z; \chi^*)|^{2-q} + \chi^* q(q-1))^3} \phi(z) dz dF(b) = 0.$$

Finally a direct application of Dominated Convergence Theorem gives us $\sigma^{2q-2} U_1 \to q^2 \mathbb{E}|B|^{2q-2}$. Hence we are able to derive the following

$$\lim_{\sigma \to 0} \frac{\chi^*}{\sigma^q} = \lim_{\sigma \to 0} \frac{\sigma^{q-2} U_2 + \sigma^{q-2} \chi^* U_3}{\sigma^{2q-2} U_1} = \frac{(q-1)\mathbb{E}|B|^{q-2}}{q \mathbb{E}|B|^{2q-2}}.$$

Now that we have derived the convergence rate of χ^*, according to Lemma 3.4.10, we can immediately obtain the order of $R_q(\chi^*, \sigma)$. □

Having the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$ as $\sigma \to 0$ in Lemma 3.4.14, the derivation for the expansion of AMSE$(\lambda_{*,q}, q, \sigma_w)$ will be the same as the one in the proof of Theorem 3.2.3.

3.4.4.2 Proof for the case $q = 1$

Lemma 3.4.15. *Suppose that $P(|B| \leq t) = \Theta(\sigma^\ell)$ (as $t \to 0$) and $\mathbb{E}|B|^2 < \infty$, then for $q = 1$*

$$\alpha_m \sigma^\ell \leq \chi_q^*(\sigma) \leq \beta_m \sigma^\ell (\log_m(1/\sigma))^{\ell/2},$$
$$\tilde{\alpha}_m \sigma^{2\ell} \leq 1 - R_q(\chi_q^*(\sigma), \sigma) \leq \tilde{\beta}_m \sigma^{2\ell} (\log_m(1/\sigma))^{\ell},$$

for small enough σ, where $\log_m(1/\sigma) = \underbrace{\log\log\ldots\log}_{m \text{ times}}\left(\frac{1}{\sigma}\right)$; $m > 0$ is an arbitrary integer number; and $\alpha_m, \beta_m, \tilde{\alpha}_m, \tilde{\beta}_m > 0$ are four constants depending on m.

Proof. Since the proof steps are similar to those in Lemma 3.4.5, we do not repeat every detail and instead highlight the differences. We write χ^* for $\chi_q^*(\sigma)$ for notational simplicity. Using the same proof steps in Lemma 3.4.5, we can obtain $\chi^* \to 0$, as $\sigma \to 0$ and

$$\chi^* = \frac{\mathbb{E}\phi(\chi^* - B/\sigma) + \mathbb{E}\phi(\chi^* + B/\sigma)}{\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*)}.$$

Following the same arguments from the proof of Lemma 2.5.34, we can show

$$\Theta(\sigma^\ell) \leq \mathbb{E}\phi(\chi^* - B/\sigma) + \mathbb{E}\phi(\chi^* + B/\sigma) \leq \Theta(\sigma^\ell (\log_m(1/\sigma))^{\ell/2}), \quad (3.40)$$
$$\Theta(\sigma^\ell) \leq \mathbb{E}\phi(\sqrt{2}B/\sigma), \quad \mathbb{E}\phi(-B/\sigma + \alpha\chi^*) \leq \Theta(\sigma^\ell (\log_m(1/\sigma))^{\ell/2}), \quad (3.41)$$

where α is any number between 0 and 1. Since $\mathbb{E}\mathbb{1}(|Z + B/\sigma| \geq \chi^*) \to 1$, the bounds for χ^* is proved by using the result (3.40). Furthermore, we know

$$R_q(\chi^*, \sigma) - 1 \leq R_q(\chi, \sigma) - 1 = \mathbb{E}(\eta_1(B/\sigma + Z; \chi) - B/\sigma - Z)^2 +$$
$$2\mathbb{E}(\partial_1 \eta_1(B/\sigma + Z; \chi) - 1) \leq \chi^2 - 2\mathbb{E}\int_{-B/\sigma - \chi}^{-B/\sigma + \chi} \phi(z)dz = \chi^2 - 4\chi\mathbb{E}\phi(-B/\sigma + \alpha\chi),$$

where $|\alpha| \leq 1$ is dependent on B. If we choose $\chi = 3e^{-1}\mathbb{E}\phi(\sqrt{2}B/\sigma)$ in the above inequality, it is straightforward to see that

$$R_q(\chi^*, \sigma) - 1 \leq -\Theta((\mathbb{E}\phi(\sqrt{2}B/\sigma))^2) \leq -\Theta(\sigma^{2\ell}),$$

where the last step is due to (3.41). For the other bound, note that

$$
\begin{aligned}
R_q(\chi^*, \sigma) - 1 &= \mathbb{E}(\eta_1(B/\sigma + Z; \chi^*) - B/\sigma - Z)^2 + 2\mathbb{E}(\partial_1 \eta_1(B/\sigma + Z; \chi^*) - 1) \\
&\geq -2\mathbb{E}\int_{-B/\sigma-\chi^*}^{-B/\sigma+\chi^*} \phi(z)dz = -4\chi^*\mathbb{E}\phi(-B/\sigma + \alpha\chi^*) \geq -\Theta(\sigma^{2\ell}(\log_m(1/\sigma))^{\ell}).
\end{aligned}
$$

The last inequality holds because of the upper bound on χ^* and (3.41). □

Based on the results of Lemma 3.4.15, deriving the expansion of $\text{AMSE}(\lambda_{*,1}, 1, \sigma_{\text{w}})$ can be done in a similar way as in the proof of Theorem 3.2.3. We do not repeat it here.

3.4.5 Proof of Theorem 3.3.2

The idea of this proof is similar to those for Theorems 3.2.3 and 3.2.5. We make use of the result in Theorem 2.2.2:

$$
\text{AMSE}(\lambda_{*,q}, q, \delta) = \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) = \delta\bar{\sigma}^2 - \sigma_w^2. \tag{3.42}
$$

Since we are in the large sample regime where $\delta \to \infty$, $\bar{\sigma}$ is a function of δ. It is clear from (3.42) that $0 \leq \delta\bar{\sigma}^2 - \sigma_w^2 \leq \bar{\sigma}^2$. Hence $\bar{\sigma}^2 \leq \sigma_w^2/(\delta - 1) \to 0$, which further leads to

$$
\bar{\sigma}^2 = \frac{\sigma_w^2}{\delta} + o(1/\delta). \tag{3.43}
$$

Due to the fact that $\bar{\sigma} \to 0$ as $\delta \to \infty$, we will be able to use the convergence rate results of $R_q(\chi_q^*(\sigma), \sigma)$ (as $\sigma \to 0$) we have proved in Lemmas 3.4.4 and 3.4.5. For $1 < q \leq 2$, Equations (3.42), (3.43) and Lemma 3.4.4 together yield

$$
\begin{aligned}
&\delta^2(\text{AMSE}(\lambda_{*,q}, q, \delta) - \sigma_w^2/\delta) = \delta^2(\bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - (\bar{\sigma}^2 - \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma})/\delta)) \\
&= (\bar{\sigma}^4\delta^2) \cdot \frac{R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1}{\bar{\sigma}^2} + (\delta\bar{\sigma}^2) \cdot R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) \\
&\to \frac{-(q-1)^2(\mathbb{E}|B|^{q-2})^2}{\mathbb{E}|B|^{2q-2}}\sigma_w^4 + \sigma_w^2.
\end{aligned} \tag{3.44}
$$

For the case $q = 1$, from Lemma 3.4.5 we know $R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1$ is exponentially small. So the firs term in (3.44) vanishes and the second term remains the same.

3.4.6 Proof of Theorem 3.3.3

Theorem 3.3.3 can be proved in a similar fashion as for Theorem 3.3.2. Equation (3.43) still holds. Equations (3.42), (3.43) and Lemma 3.4.15 together give us for $q = 1$,

$$\delta^{\ell+1}(\text{AMSE}(\lambda_{*,q}, q, \delta) - \sigma_w^2/\delta) = (\bar{\sigma}^{2\ell+2}\delta^{\ell+1}) \cdot \frac{R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - 1}{\bar{\sigma}^{2\ell}} + (\delta^{\ell}\bar{\sigma}^2) \cdot R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}),$$

where the first term above is $\Theta(1)$ and the second one is $o(1)$ when $\ell < 1$. The case $1 < q \leq 2$ can be proved exactly the same way as in Theorem 3.3.2 by using Lemma 3.4.14.

Chapter 4

From low noise to large noise analysis

4.1 Introduction

We have presented a thorough analysis of bridge estimators under the low noise regime in Chapters 2 and 3. Our analysis can be considered as a generalization of phase transition analysis and provides a more accurate characterization of LQLS estimators when a small noise is added in the model. Nevertheless, in many applications the noise present in the data is very large. Hence our preceding analysis can be irrelevant in such cases. In this chapter, we adopt the same sensitivity analysis framework we used in the previous two chapters, and characterize the performance of LQLS in low signal-to-noise ratios. This time we let the noise level $\sigma_w \to \infty$ and derive the second-order expansion of AMSE. Our results reveal a completely different picture of the behavior of LQLS estimators from what we have showed in the low noise setting. In particular, among all the $q \in [1, \infty)$, ridge is optimal and LASSO is always suboptimal.

4.2 A second-order large noise sensitivity analysis

Recall that G is the random variable that characterizes the non-zero components of the coefficient β, and we use Z to denote a standard normal. The main result is presented in the theorem below[1].

Theorem 4.2.1. *As $\sigma \to \infty$, we have the following expansions of* $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$.

(i) For $q = 1$, when G has sub-Gaussian tail, we have

$$\mathrm{AMSE}(\lambda_{*,1}, 1, \sigma_w) = \epsilon \mathbb{E}|G|^2 + o(e^{-\frac{C^2\sigma_w^2}{2}}), \tag{4.1}$$

where C is any positive number smaller than C_0, with C_0 a constant only depending on ϵ and G.[2]

(ii) For $1 < q \le 2$, if all the moments of G are finite, then

$$\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) = \epsilon \mathbb{E}|G|^2 - \frac{\epsilon^2(\mathbb{E}|G|^2)^2 c_q}{\sigma_w^2} + o(\sigma_w^{-2}), \tag{4.2}$$

with $c_q = \frac{\left(\mathbb{E}|Z|^{\frac{2-q}{q-1}}\right)^2}{(q-1)^2\mathbb{E}|Z|^{\frac{2}{q-1}}}$.

(iii) For $q > 2$, if G has sub-Gaussian tail, then (4.2) holds.

The proof is presented in Chapter 4.3. Figure 4.1 compares the accuracy of the first-order approximation and second-order approximation for moderate values of σ_w. As is clear, for $q \in (1, \infty)$ the second-order approximation provides a more accurate approximation for a wide range of σ_w. Moreover, the first-order approximation of LASSO is already very accurate as can be justified by its exponentially small second order term in (4.1).

According to this theorem, we can conclude that for sufficiently large σ_w, LQLS with any $q > 1$ can outperform LASSO. This is because while the first dominant term

[1]The result is part of Wang et al. (2017).

[2]Refer to the proof of this theorem for the exact characterization of C_0.

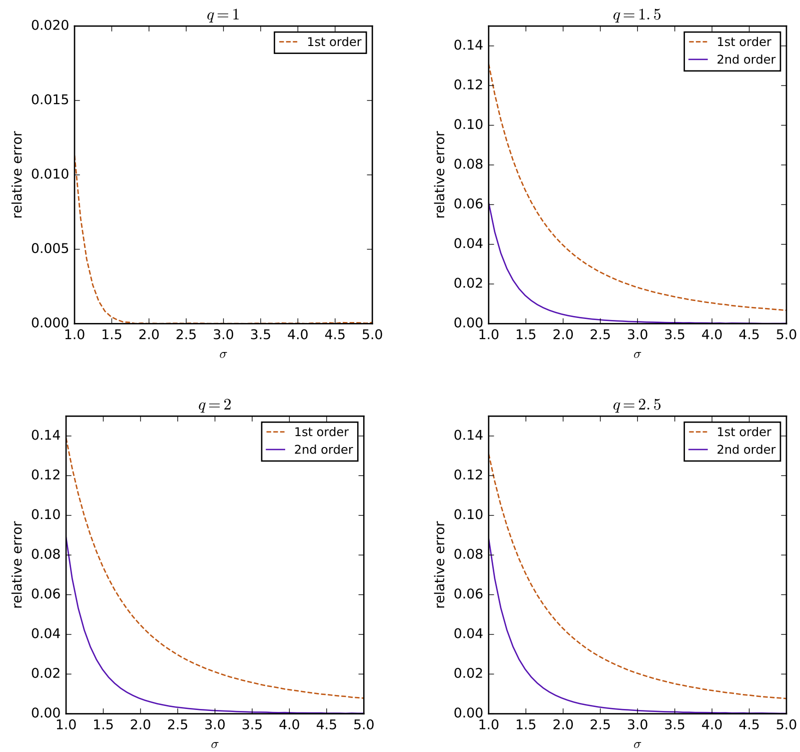


Figure 4.1: Absolute relative error of first-order and second-order approximations of AMSE, displayed by the orange curves and purple curves respectively, for different values of σ_w. In these four figures, $p_B = (1-\epsilon)\delta_0 + \epsilon\delta_1$, $\delta = 0.4$, $\epsilon = 0.2$.

is the same for all the bridge estimators with $q \in [1, \infty)$, the second order term for LASSO is exponentially smaller (in magnitude) than that of the other values of q. More interestingly, the following corollary shows that in fact $q = 2$ leads to the best AMSE in the large noise regime.

Corollary 4.2.2. *The maximum of* c_q*, defined in Theorem 4.2.1, is achieved at* $q = 2$.

Proof. A simple integration by part yields:

$$\begin{aligned} \mathbb{E}|Z|^{\frac{2-q}{q-1}} &= 2\int_0^\infty z^{\frac{2-q}{q-1}}\phi(z)dz = 2(q-1)\int_0^\infty \phi(z)dz^{\frac{1}{q-1}} \\ &= 2(q-1)\int_0^\infty z^{\frac{q}{q-1}}\phi(z)dz = (q-1)\mathbb{E}|Z|^{\frac{q}{q-1}} \end{aligned}$$

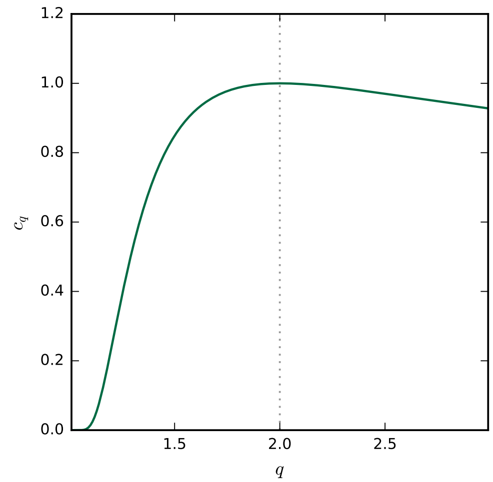


Figure 4.2: The constant c_q in Theorem 4.2.1 part (ii). The maximum is achieved at $q = 2$.

We can then apply Hölders's inequality to obtain

$$c_q = \frac{(\mathbb{E}|Z|^{\frac{q}{q-1}})^2}{\mathbb{E}|Z|^{\frac{2}{q-1}}} \leq \frac{\mathbb{E}|Z|^{\frac{2}{q-1}}\mathbb{E}Z^2}{\mathbb{E}|Z|^{\frac{2}{q-1}}} = 1 = c_2.$$

□

Therefore while the AMSE of all bridge estimators share the same first dominant term, ridge offers the largest second dominant term (in magnitude), and hence the lowest AMSE. Figure 4.3 shows the AMSE comparison for different LQLS estimators. It is important to note that the optimality of ridge and suboptimality of LASSO hold for sparse coefficients. In other words, in low signal-to-noise ratios ℓ_2-regularization gives better estimates for sparse parameters even than the sparsity promoting ℓ_1-regularization. This is in contrast to the conclusion we obtained for low noise setting. The results in low and large noises together can be considered as a manifest of bias-variance trade-off principle in statistics.

A natural question is the comparison among $q \in [0, 1]$. Theorem 4.2.1 may imply that $q = 1$ can outperform all the other $q < 1$, since ℓ_q-regularization with $q < 1$ is

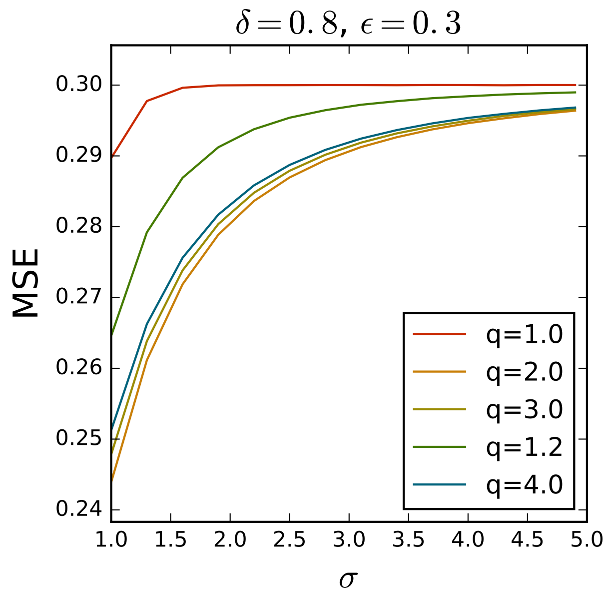


Figure 4.3: The AMSE curves for LQLS estimators with $q = 1, 1.2, 2, 3, 4$. For this figure, $p_B = (1-\epsilon)\delta_0 + \epsilon\delta_1$, $\delta = 0.8$, $\epsilon = 0.3$.

even more aggressive than ℓ_1 hence leading to larger variance. We formally confirm it for a special family of distributions f_β in the next theorem.

Theorem 4.2.3. *Suppose $f_\beta = (1-\epsilon)\delta_0 + \epsilon\delta_\mu$, where μ is a non-zero constant. Then for any given $0 \leq q < 1$, there exists a threshold $\bar{\sigma}_w$ such that*

$$\text{AMSE}(\lambda_{*,1}, 1, \sigma_w) < \text{AMSE}(\lambda_{*,q}, q, \sigma_w), \quad \textit{for all } \sigma_w > \bar{\sigma}_w.$$

The proof can be found in Zheng et al. (2017).

4.3 Proof of Theorem 4.2.1

4.3.1 Roadmap and preliminaries

Since the proof of this theorem is long, we lay out the roadmap of the proof here to help readers navigate through the details. According to Corollary 2.2.4, we know that $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ can be computed as

$$\text{AMSE}(\lambda_{*,q}, q, \sigma_w) = \delta(\bar{\sigma}^2 - \sigma_w^2), \tag{4.3}$$

where $\bar{\sigma}$ satisfies the following equation:

$$\bar{\sigma}^2 = \sigma_w^2 + \frac{1}{\delta} \min_{\alpha \geq 0} \mathbb{E}(\eta_q(B + \bar{\sigma} Z; \alpha \bar{\sigma}^{2-q}) - B)^2. \tag{4.4}$$

It is clear from (4.4) that $\bar{\sigma} \to \infty$ as $\sigma_w \to \infty$. However, to derive the second-order expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ as $\sigma_w \to \infty$, we need to obtain the convergence rate of $\bar{\sigma}$. We will achieve this goal by first characterizing the convergence rate of the term $\min_{\alpha \geq 0} \mathbb{E}(\eta_q(B + \sigma Z; \alpha \sigma^{2-q}) - B)^2$ as $\sigma \to \infty$. We then use that result to derive the convergence rate of $\bar{\sigma}$ based on (4.4) and finally calculate $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ through (4.3). Since the proof techniques look different for $q = 1, 1 < q \leq 2, q > 2$, we prove the theorem for these three cases separately.

Let Φ and ϕ denote the cumulative distribution function and probability density function of a standard normal random variable respectively. Standard result on the expansion of Gaussian tails through integration by parts gives: for $k \in \mathbb{N}^+, s > 0$

$$\Phi(-s) = \phi(s) \Big[\sum_{i=0}^{k-1} \frac{(-1)^i (2i-1)!!}{s^{2i+1}} + (-1)^k (2k-1)!! \int_s^\infty \frac{\phi(t)}{t^{2k}} dt \Big], \tag{4.5}$$

where $(2i-1)!! \triangleq 1 \times 3 \times 5 \times \ldots \times (2i-1)$.

Lemma 4.3.1. *Consider a nonnegative random variable X with probability distribution μ and $\mathbb{P}(X > 0) = 1$. Let $\xi > \zeta > 0$ be the points such that $\mathbb{P}(X \leq \zeta) \leq \frac{1}{4}$ and $\mathbb{P}(\zeta < X \leq \xi) \geq \frac{1}{4}$. Let $a, b, c : \mathbb{R}_+ \to \mathbb{R}_+$ be three deterministic positive functions such that $a(s), c(s) \to \infty$ as $s \to \infty$. Then there exists a positive constant s_0 depending on a, c, X, such that when $s > s_0$,*

$$\int_0^{a(s)} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x) \leq 3 \int_\zeta^{a(s)} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x).$$

Proof. For large enough s such that $a(s) > \xi$,

$$\int_{\zeta}^{a(s)} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x) \geq \int_{\zeta}^{\xi} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x) \geq e^{b(s)\zeta - \frac{\xi^2}{c(s)}} \mathbb{P}(\zeta < X \leq \xi)$$
$$\geq e^{b(s)\zeta - \frac{\xi^2}{c(s)}} \mathbb{P}(X \leq \zeta) \geq e^{-\frac{\xi^2}{c(s)}} \int_{0}^{\zeta} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x).$$

As a result we have the following inequality,

$$\int_{0}^{a(s)} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x) \leq (1 + e^{\frac{\xi^2}{c(s)}}) \int_{\zeta}^{a(s)} e^{b(s)x - \frac{x^2}{c(s)}} d\mu(x).$$

For sufficiently large s such that $e^{\frac{\xi^2}{c(s)}} < 2$, the conclusion follows. □

4.3.2 Proof of Theorem 4.2.1 for $q = 1$

According to Definitions (2.46), (2.77) and Lemma 2.5.5 part (v), it is clear that Equation (4.4) can be rewritten:

$$\bar{\sigma}^2 = \sigma_w^2 + \frac{1}{\delta} \bar{\sigma}^2 R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}). \tag{4.6}$$

As explained in the roadmap of the proof, the key step is to characterize the convergence rate of $\bar{\sigma}$. Towards this goal, we first derive the convergence rate of $\chi_q^*(\sigma)$ as $\sigma \to \infty$ in Chapter 4.3.2.1. We then bound the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$ as $\sigma \to \infty$ in Chapter 4.3.2.2. We finally apply the preceding result to (4.6) to characterize $\bar{\sigma}$ when $\sigma_w \to \infty$, and derive the expansion of $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$ as $\sigma_w \to \infty$ in Chapter 4.3.2.3.

4.3.2.1 Deriving the convergence rate of $\chi_q^*(\sigma)$ as $\sigma \to \infty$ for $q = 1$

We first prove $\chi_q^*(\sigma) \to \infty$ as $\sigma \to \infty$ in the next lemma.

Lemma 4.3.2. *Assume* $\mathbb{E}|G|^2 < \infty$. *Then,* $\chi_q^*(\sigma) \to \infty$ *as* $\sigma \to \infty$.

Proof. Suppose this is not true, then there exists a sequence $\{\sigma_n\}$ such that $\chi_q^*(\sigma_n) \to \chi_0 < \infty$ and $\sigma_n \to \infty$, as $n \to \infty$. Notice that

$$|\eta_q(B/\sigma_n + Z; \chi_q^*(\sigma_n))| \leq |B|/\sigma_n + Z \leq |B| + Z,$$

for large n. We can apply Dominated Convergence Theorem (DCT) to obtain

$$\lim_{n\to\infty} R_q(\chi_q^*(\sigma_n), \sigma_n) = \mathbb{E}\eta_q^2(Z; \chi_0) > 0.$$

On the other hand, since $\chi = \chi_q^*(\sigma_n)$ minimizes $R_q(\chi, \sigma_n)$

$$\lim_{n\to\infty} R_q(\chi_q^*(\sigma_n), \sigma_n) \leq \lim_{n\to\infty}\lim_{\chi\to\infty} R_q(\chi, \sigma_n) = 0.$$

A contradiction arises. □

Based on Lemma 4.3.2, we can further derive the convergence rate of $\chi_q^*(\sigma)$.

Lemma 4.3.3. *If G has a sub-Gaussian tail, then*

$$\lim_{\sigma\to\infty} \frac{\chi_q^*(\sigma)}{\sigma} = C_0,$$

where $C = C_0$ is the unique solution of the following equation:

$$\mathbb{E}\left(e^{CG}(CG-1) + e^{-CG}(-CG-1)\right) = \frac{2(1-\epsilon)}{\epsilon}.$$

Proof. Since $\chi = \chi_q^*(\sigma)$ minimizes $R_q(\chi, \sigma)$, we know

$$\partial_1 R_q(\chi_q^*(\sigma), \sigma) = 0. \tag{4.7}$$

To simplify the notation, we will simply write χ for $\chi_q^*(\sigma)$ in the rest of this proof. Rearranging the terms in (4.7) gives us

$$\frac{2(1-\epsilon)}{\epsilon} = \mathbb{E}\underbrace{\frac{\chi^2}{\phi(\chi)}\Big[\chi\Phi\big(\frac{|G|}{\sigma}-\chi\big) + \chi\Phi\big(-\frac{|G|}{\sigma}-\chi\big) - \phi\big(\frac{|G|}{\sigma}-\chi\big) - \phi\big(\frac{|G|}{\sigma}+\chi\big)\Big]}_{T(G,\chi,\sigma)}.$$

Fixing $t \in (0,1)$, we reformulate the above equation in the following way:

$$\frac{2(1-\epsilon)}{\epsilon} = \mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| \leq t\sigma\chi)] + \mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| > t\sigma\chi)]. \tag{4.8}$$

We now analyze the two terms on the right hand side of the above equation. Since G has a sub-Gaussian tail, there exists a constant $\gamma > 0$ such that $\mathbb{P}(|G| > x) \leq e^{-\gamma x^2}$ for x large. We can then have the following bound,

$$\begin{aligned}
&|\mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| > t\sigma\chi)]| \leq \frac{\chi^2}{\phi(\chi)}\big(2\chi + \sqrt{2/\pi}\big)\mathbb{P}(|G| > t\sigma\chi) \\
&\leq \chi^2\big(2\sqrt{2\pi}\chi + 2\big)e^{-(\gamma t^2\sigma^2 - \frac{1}{2})\chi^2} \to 0, \quad \text{as } \sigma\to\infty,
\end{aligned}$$

where we have used the fact that $\chi \to \infty$ as $\sigma \to \infty$ from Lemma 4.3.2. This result combined with (4.8) implies that as $\sigma \to \infty$

$$\mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| \le t\sigma\chi)] \to \frac{2(1-\epsilon)}{\epsilon}. \tag{4.9}$$

Moreover, using the tail approximation of normal distribution in (4.5) with $k = 3$, we have for sufficiently large σ,

$$\begin{aligned}
&\mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| \le t\sigma\chi)] \\
&\le \mathbb{E}\Big[\underbrace{\frac{\chi}{\chi - |G|/\sigma} e^{\frac{\chi|G|}{\sigma} - \frac{G^2}{2\sigma^2}} \left(\frac{\chi|G|}{\sigma} - \frac{\chi^2}{(\chi - |G|/\sigma)^2} + \frac{3\chi^2}{(\chi - |G|/\sigma)^4}\right)}_{U_1(G,\chi,\sigma)} + \\
&\underbrace{\frac{\chi}{\chi + |G|/\sigma} e^{-\frac{\chi|G|}{\sigma} - \frac{G^2}{2\sigma^2}} \left(-\frac{\chi|G|}{\sigma} - \frac{\chi^2}{(\chi + |G|/\sigma)^2} + \frac{3\chi^2}{(\chi + |G|/\sigma)^4}\right)}_{U_2(G,\chi,\sigma)}\Big] \cdot \mathbb{I}(|G| \le t\sigma\chi).
\end{aligned}$$

Similarly applying (4.5) with $k = 2$ gives us for large σ

$$\begin{aligned}
\mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G| \le t\sigma\chi)] &\ge \mathbb{E}\Big[\underbrace{\frac{\chi}{\chi - |G|/\sigma} e^{\frac{\chi|G|}{\sigma} - \frac{G^2}{2\sigma^2}} \left(\frac{\chi|G|}{\sigma} - \frac{\chi^2}{(\chi - |G|/\sigma)^2}\right)}_{L_1(G,\chi,\sigma)} + \\
&\underbrace{\frac{\chi}{\chi + |G|/\sigma} e^{-\frac{\chi|G|}{\sigma} - \frac{G^2}{2\sigma^2}} \left(-\frac{\chi|G|}{\sigma} - \frac{\chi^2}{(\chi + |G|/\sigma)^2}\right)}_{L_2(G,\chi,\sigma)}\Big] \cdot \mathbb{I}(|G| \le t\sigma\chi).
\end{aligned}$$

We can conclude based on the two bounds that $\overline{\lim}_{\sigma \to \infty} \frac{\chi}{\sigma} = C_1$ with $0 < C_1 < \infty$. Otherwise

- If $C_1 = \infty$, there exists a sequence $\chi_n/\sigma_n \to \infty$ and $\sigma_n \to \infty$, as $n \to \infty$. Since $|L_2(G,\chi_n,\sigma_n)| \le e^{-\frac{\chi_n|G|}{\sigma_n}} (\frac{\chi_n|G|}{\sigma_n} + 1) \le 2$, we can apply DCT to obtain

$$\lim_{n \to \infty} \mathbb{E}(L_2(G,\chi_n,\sigma_n)\mathbb{I}(|G| \le t\sigma_n\chi_n)) = 0.$$

Furthermore, we choose a positive constant $\zeta > 0$ satisfying the condition in

Lemma 4.3.1 for the nonnegative random variable $|G|$. Then

$$\begin{aligned}
&\mathbb{E}(L_1(G,\chi_n,\sigma_n)\mathbb{I}(|G|\le t\sigma_n\chi_n)) \\
&\ge \mathbb{E}\Big[e^{\frac{\chi_n|G|}{\sigma_n}-\frac{G^2}{2\sigma_n^2}}\Big(\frac{\chi_n|G|}{\sigma_n}-\frac{1}{(1-t)^3}\Big)\mathbb{I}(|G|\le t\sigma_n\chi_n)\Big] \\
&\ge \int_{\zeta<g\le t\sigma_n\chi_n} e^{\frac{\chi_n g}{\sigma_n}-\frac{g^2}{2\sigma_n^2}}\frac{\chi_n g}{\sigma_n}dF(g) - \int_{g\le t\sigma_n\chi_n}\frac{1}{(1-t)^3}e^{\frac{\chi_n g}{\sigma_n}-\frac{g^2}{2\sigma_n^2}}dF(g) \\
&\overset{(a)}{\ge} \Big(\frac{\zeta\chi_n}{\sigma_n}-\frac{2}{(1-t)^3}\Big)\int_{\zeta<g\le t\sigma_n\chi_n} e^{\frac{\chi_n g}{\sigma_n}-\frac{g^2}{2\sigma_n^2}}dF(g) \\
&\ge \Big(\frac{\zeta\chi_n}{\sigma_n}-\frac{2}{(1-t)^3}\Big)e^{\frac{\chi_n\zeta}{\sigma_n}}\int_{\zeta<g\le t\sigma_n\chi_n} e^{-\frac{g^2}{2\sigma_n^2}}dF(g)\to\infty,
\end{aligned}$$

where we have used Lemma 4.3.1 in (a). This forms a contradiction.

- If $C_1=0$, for large enough σ we have $\frac{\chi}{\sigma}<1$ and then on $|G|\le t\sigma\chi$,

$$|U_1(G,\chi,\sigma)+U_2(G,\chi,\sigma)|\le\frac{2}{1-t}e^G\Big[G+\frac{1}{(1-t)^2}+\frac{3}{\chi^2(1-t)^4}\Big],$$

which is integrable since G has sub-Gaussian tail. Hence we apply DCT to obtain as $\sigma\to\infty$

$$\mathbb{E}\left[(U_1(G,\chi,\sigma)+U_2(G,\chi,\sigma))\mathbb{I}(|G|\le t\sigma\chi)\right]\to -2$$

This forms another contradiction.

Similar to the above arguments, we can conclude that $\underline{\lim}_{\sigma\to\infty}\frac{\chi}{\sigma}=C_2\in(0,\infty)$. Now that $\frac{\chi}{\sigma}=O(1)$, we can use DCT to obtain

$$\lim_{\sigma\to\infty}\mathbb{E}\Big[\frac{\chi}{\chi\pm|G|/\sigma}e^{\frac{\chi|G|}{\sigma}-\frac{G^2}{2\sigma^2}}\frac{3\chi^2}{(\chi\pm|G|/\sigma)^4}\mathbb{I}(|G|\le t\sigma\chi)\Big]=0.$$

This result combined with (4.9) and the upper and lower bounds on $\mathbb{E}[T(G,\chi,\sigma)\mathbb{I}(|G|\le t\sigma\chi)]$ enables us to show

$$\lim_{\sigma\to\infty}\mathbb{E}[(L_1(G,\chi,\sigma)+L_2(G,\chi,\sigma))\mathbb{I}(|G|\le t\sigma\chi)]=\frac{2(1-\epsilon)}{\epsilon}.$$

Now consider a convergent sequence $\frac{\chi_n}{\sigma_n}\to C_1\in(0,\infty)$ and $\sigma_n\to\infty$ as $n\to\infty$. On $|G|\le t\sigma_n\chi_n$ we can bound for large n

$$|L_1(G,\chi_n,\sigma_n)+L_2(G,\chi_n,\sigma_n)|\le\frac{2}{1-t}e^{2C_1G}\Big(2C_1G+\frac{1}{(1-t)^2}\Big),$$

which is again integrable. Thus DCT gives us

$$\begin{aligned}\frac{2(1-\epsilon)}{\epsilon} &= \lim_{n\to\infty} \mathbb{E}[(L_1(G,\chi_n,\sigma_n)+L_2(G,\chi_n,\sigma_n))\mathbb{I}(|G|\le t\sigma_n\chi_n)] \\ &= \mathbb{E}\big[e^{C_1|G|}(C_1|G|-1)+e^{-C_1|G|}(-C_1|G|-1)\big].\end{aligned}$$

For C_2 the same equation holds. By calculating the derivative we can easily verify $h(c)=e^{c|G|}(c|G|-1)+e^{-c|G|}(-c|G|-1)$, as a function of c over $(0,\infty)$, is strictly increasing. This determines $C_1=C_2$. Above all we have shown

$$\frac{\chi_q^*(\sigma)}{\sigma}\to C_0, \quad \text{as } \sigma\to\infty,$$

where $\mathbb{E}\big[e^{C_0G}(C_0G-1)+e^{-C_0G}(-C_0G-1)\big]=\frac{2(1-\epsilon)}{\epsilon}$. □

4.3.2.2 Bounding the convergence rate of $R_q(\chi_q^*(\sigma),\sigma)$ as $\sigma\to\infty$ for $q=1$

We state the main result in the next lemma.

Lemma 4.3.4. *If G has sub-Gaussian tail, then as $\sigma\to\infty$*

$$R_q(\chi_q^*(\sigma),\sigma)=\frac{\epsilon\mathbb{E}|G|^2}{\sigma^2}+o\Big(\frac{\phi(\chi_q^*(\sigma))}{(\chi_q^*(\sigma))^3}\Big).$$

Proof. For notational simplicity, we will use χ to denote $\chi_q^*(\sigma)$ in the rest of the proof. Since $\eta_1(u;\chi)=\text{sgn}(u)(|u|-\chi)_+$, we can write $R_q(\chi,\sigma)$ in the following form:

$$\begin{aligned}R_q(\chi,\sigma) &= 2(1-\epsilon)\big((1+\chi^2)\Phi(-\chi)-\chi\phi(\chi)\big) \\ &+\epsilon\mathbb{E}\Bigg[\underbrace{(1+\chi^2-G^2/\sigma^2)\Big(\Phi(G/\sigma-\chi)+\Phi(-G/\sigma-\chi)\Big)}_{S_1(G,\chi,\sigma)} \\ &\underbrace{-(G/\sigma+\chi)\phi(\chi-G/\sigma)+(G/\sigma-\chi)\phi(\chi+G/\sigma)}_{S_2(G,\chi,\sigma)}+G^2/\sigma^2\Bigg].\end{aligned}$$

Hence, we have

$$\begin{aligned}\lim_{\sigma\to\infty}\frac{\chi^3}{\phi(\chi)}\Big(R_q(\chi,\sigma)-\frac{\epsilon\mathbb{E}|G|^2}{\sigma^2}\Big) &= 2(1-\epsilon)\lim_{\sigma\to\infty}\frac{\chi^3}{\phi(\chi)}[(1+\chi^2)\Phi(-\chi)-\chi\phi(\chi)] \\ &+\epsilon\lim_{\sigma\to\infty}\frac{\chi^3}{\phi(\chi)}\mathbb{E}[S_1(G,\chi,\sigma)+S_2(G,\chi,\sigma)] \\ &\overset{(a)}{=} 4(1-\epsilon)+\epsilon\lim_{\sigma\to\infty}\frac{\chi^3}{\phi(\chi)}\mathbb{E}[S_1(G,\chi,\sigma)+S_2(G,\chi,\sigma)]. \qquad (4.10)\end{aligned}$$

We have used the tail expansion (4.5) with $k = 3, 4$ to obtain (a). Note that since $|x\phi(x)| \leq \frac{e^{-1/2}}{\sqrt{2\pi}}$, we have

$$|S_1(G,\chi,\sigma) + S_2(G,\chi,\sigma)| \leq \frac{2e^{-1/2}}{\sqrt{2\pi}} + \frac{4\chi}{\sqrt{2\pi}} + 2\bigg(1 + \chi^2 + \frac{G^2}{\sigma^2}\bigg).$$

Moreover, it is not hard to use the sub-Gaussian condition $\mathbb{P}(|G| > x) \leq e^{-\gamma x^2}$ to obtain

$$\begin{aligned}\mathbb{E}(G^2\mathbb{I}(|G| > t\sigma\chi)) &= \int_0^{t\sigma\chi} 2x\mathbb{P}(G > t\sigma\chi)dx + \int_{t\sigma\chi}^{\infty} 2x\mathbb{P}(G > x)dx \\ &\leq (t\sigma\chi)^2 e^{-\gamma t^2\sigma^2\chi^2} + \frac{1}{\gamma}e^{-\gamma t^2\sigma^2\chi^2},\end{aligned}$$

where $t \in (0,1)$ is a constant. Combining the last two bounds we can derive

$$\begin{aligned}&\frac{\chi^3}{\phi(\chi)}\mathbb{E}[(S_1(G,\chi,\sigma) + S_2(G,\chi,\sigma))\mathbb{I}(|G| > t\sigma\chi)] \\ \leq\ &\chi^3\big(2e^{-1/2} + 4\chi + 2\sqrt{2\pi}(1+\chi^2)\big)e^{-(\gamma t^2\sigma^2 - \frac{1}{2})\chi^2} + \\ &\frac{2\sqrt{2\pi}\chi^3}{\sigma^2}(t^2\sigma^2\chi^2 + 1/\gamma)e^{-(\gamma t^2\sigma^2 - \frac{1}{2})\chi^2} \to 0, \quad \text{as } \sigma \to \infty.\end{aligned}$$

On the other hand, we can build an upper bound and lower bound for $|S_1(G,\chi,\sigma) + S_2(G,\chi,\sigma)|$ on $\{|G| \leq t\sigma\chi\}$ with the tail expansion (4.5) as we did in the proof of Lemma 4.3.3, For both bounds we can argue they converge to the same limit as $\sigma \to \infty$ by using DCT and Lemma 4.3.3. Here we give the details of using DCT for the upper bound. Using (4.5) with $k = 3$ we can obtain the upper bound,

$$\begin{aligned}&\frac{\chi^3}{\phi(\chi)}(S_1(G,\chi,\sigma) + S_2(G,\chi,\sigma)) \\ \leq\ &\frac{\chi^3\phi(\chi - G/\sigma)}{\phi(\chi)}\left[\frac{2G^2/\sigma^2 - 2\chi G/\sigma - 1}{(\chi - G/\sigma)^3} + \frac{3(1+\chi^2 - G^2/\sigma^2)}{(\chi - G/\sigma)^5}\right] + \\ &\frac{\chi^3\phi(\chi + G/\sigma)}{\phi(\chi)}\left[\frac{2G^2/\sigma^2 + 2\chi G/\sigma - 1}{(\chi + G/\sigma)^3} + \frac{3(1+\chi^2 - G^2/\sigma^2)}{(\chi + G/\sigma)^5}\right].\end{aligned}$$

It is straightforward to see that on $\{|G| \leq t\sigma\chi\}$ for sufficiently large χ, there exist three positive constants C_1, C_2, C_3 such that the upper bound can be further bounded by $\left[\frac{C_1|G|+1}{(1-t)^3} + \frac{C_2}{(1-t)^5} + \frac{C_1|G|+1}{(1+t)^3} + \frac{C_2}{(1+t)^5}\right]e^{C_3|G|}$, which is integrable by the condition that

G has sub-Gaussian tail. Hence we can apply DCT to derive the limit of the upper bound. Similar arguments enable us to calculate the limit of the lower bound. By calculating the limits of the upper and lower bounds we can obtain the following result:

$$\begin{aligned}&\frac{\chi^3}{\phi(\chi)}\mathbb{E}[(S_1(G,\chi,\sigma)+S_2(G,\chi,\sigma))\mathbb{I}(|G|\leq t\sigma\chi)]\\&\to -2\mathbb{E}\left(e^{C_0G}(C_0G-1)+e^{-C_0G}(-C_0G-1)\right)=-\frac{4(1-\epsilon)}{\epsilon}.\end{aligned}$$

This completes the proof. □

4.3.2.3 Deriving the expansion of $\mathrm{AMSE}(\lambda_{*,q},q,\sigma_w)$ for $q=1$

We are now in the position to derive the result (4.1) in Theorem 4.2.1. As we explained in the roadmap, we know

$$\mathrm{AMSE}(\lambda_{*,q},q,\sigma_w)=\bar{\sigma}^2R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})=\delta(\bar{\sigma}^2-\sigma_w^2). \tag{4.11}$$

First note that $\bar{\sigma}\to\infty$ as $\sigma_w\to\infty$ since $\bar{\sigma}\geq\sigma_w$. Then according to Lemma 4.3.4 and (4.11), we have

$$\lim_{\sigma_w\to\infty}\frac{\sigma_w^2}{\bar{\sigma}^2}=\lim_{\bar{\sigma}\to\infty}\frac{\sigma_w^2}{\bar{\sigma}^2}=\lim_{\bar{\sigma}\to\infty}\big(1-\frac{R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})}{\delta}\big)=1. \tag{4.12}$$

Furthermore, Lemma 4.3.3 shows that

$$\lim_{\sigma_w\to\infty}\frac{\chi_q^*(\bar{\sigma})}{\bar{\sigma}}=\lim_{\bar{\sigma}\to\infty}\frac{\chi_q^*(\bar{\sigma})}{\bar{\sigma}}=C_0. \tag{4.13}$$

Combining Lemma 4.3.4 with (4.11), (4.12), and (4.13) we obtain as $\sigma_w\to\infty$,

$$\begin{aligned}&e^{\frac{C^2\sigma_w^2}{2}}(\mathrm{AMSE}(\lambda_{*,q},q,\sigma_w)-\epsilon\mathbb{E}|G|^2)=e^{\frac{C^2\sigma_w^2}{2}}\bar{\sigma}^2(R_q(\chi_q^*(\bar{\sigma}),\bar{\sigma})-\epsilon\mathbb{E}|G|^2/\bar{\sigma}^2)\\=\;&e^{\frac{C^2\sigma_w^2}{2}}\bar{\sigma}^2e^{-\frac{(\chi_q^*(\bar{\sigma}))^2}{2}}(\chi_q^*(\bar{\sigma}))^{-3}o(1)=e^{-\frac{\sigma_w^2}{2}(\frac{(\chi_q^*(\bar{\sigma}))^2}{\bar{\sigma}^2}\cdot\frac{\bar{\sigma}^2}{\sigma_w^2}-C^2)}\bar{\sigma}^2(\chi_q^*(\bar{\sigma}))^{-3}o(1)=o(1).\end{aligned}$$

We have used the fact $0<C<C_0$ to get the last equality.

4.3.3 Proof of Theorem 4.2.1 for $1 < q \le 2$

The basic idea of the proof for $q \in (1,2]$ is the same as that for $q = 1$. We characterize the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$ in Chapter 4.3.3.1. We then derive the expansion of $\text{AMSE}(\lambda_{*,q}, q, \sigma_w)$ in Chapter 4.3.3.2.

4.3.3.1 Characterizing the convergence rate of $R_q(\chi_q^*(\sigma), \sigma)$ as $\sigma \to \infty$ for $q \in (1,2]$

We first derive the convergence rate of $\chi_q^*(\sigma)$ as $\sigma \to \infty$.

Lemma 4.3.5. *For $q \in (1,2]$, assume G has finite moments of all order. We have as $\sigma \to \infty$,*

$$\frac{\chi_q^*(\sigma)}{\sigma^{2(q-1)}} \to \left(\frac{q-1}{q^{\frac{1}{q-1}}} \frac{\mathbb{E}|Z|^{\frac{2}{q-1}}}{\mathbb{E}B^2 \mathbb{E}|Z|^{\frac{2-q}{q-1}}} \right)^{q-1}.$$

Proof. First note that Lemma 4.3.2 holds for $q \in (1,2]$ as well. Hence $\chi_q^*(\sigma) \to \infty$ as $\sigma \to \infty$. We aim to characterize its convergence rate. Since $\eta_2(u;\chi) = \frac{u}{1+2\chi}$, the result can be easily verified for $q = 2$. We will focus on the case $q \in (1,2)$. For notational simplicity, we will use χ to represent $\chi_q^*(\sigma)$ in the rest of the proof. By the first order condition of the optimality, we have $\partial_1 R_q(\chi, \sigma) = 0$, which can be further written out:

$$\begin{aligned} 0 &= \mathbb{E}[(\eta_q(B/\sigma + Z;\chi) - B/\sigma)\partial_2 \eta_q(B/\sigma + Z;\chi)] \\ &= \underbrace{\mathbb{E}\left[\frac{-q|\eta_q(B/\sigma + Z;\chi)|^q}{1 + \chi q(q-1)|\eta_q(B/\sigma + Z;\chi)|^{q-2}}\right]}_{H_1} + \\ &\quad \underbrace{\frac{1}{\sigma}\mathbb{E}\left[\frac{Bq|\eta_q(B/\sigma + Z;\chi)|^{q-1}\text{sgn}(B/\sigma + Z)}{1 + \chi q(q-1)|\eta_q(B/\sigma + Z;\chi)|^{q-2}}\right]}_{H_2}, \end{aligned} \tag{4.14}$$

where we have used Lemma 2.5.8 part (ii). We now analyze the two terms H_1 and

H_2 respectively. Regarding H_1 from Lemma 2.5.5 part (i) we have

$$\frac{\chi^{\frac{q+1}{q-1}} q\big|\eta_q\big(B/\sigma + Z;\chi\big)\big|^q}{1+\chi q(q-1)\big|\eta_q\big(B/\sigma + Z;\chi\big)\big|^{q-2}} \leq \frac{\big|\chi^{\frac{1}{q-1}}\eta_q\big(B/\sigma + Z;\chi\big)\big|^2}{q-1}$$
$$= \frac{\Big||B/\sigma + Z| - \big|\eta_q(B/\sigma + Z;\chi)\big|\Big|^{\frac{2}{q-1}}}{q^{\frac{2}{q-1}}(q-1)} \leq \frac{(|B|+|Z|)^{\frac{2}{q-1}}}{q^{\frac{2}{q-1}}(q-1)}, \quad \text{for } \sigma \geq 1.$$

Since G has finite moments of all orders, the upper bound above is integrable. Hence DCT enables us to conclude

$$\lim_{\sigma\to\infty} \chi^{\frac{q+1}{q-1}} H_1 = \frac{\mathbb{E}|Z|^{\frac{2}{q-1}}}{q^{\frac{2}{q-1}}(1-q)}. \tag{4.15}$$

For the term H_2, according to Lemma 2.5.5 part (i) and Lemma 2.5.8 part (i) we can obtain

$$H_2 = \frac{1}{\sigma\chi}\mathbb{E}\left[\frac{B\big(B/\sigma + Z - \eta_q(B/\sigma + Z;\chi)\big)}{1+\chi q(q-1)\big|\eta_q\big(B/\sigma + Z;\chi\big)\big|^{q-2}}\right]$$
$$= \underbrace{\frac{1}{\sigma^2\chi}\mathbb{E}\left[\frac{B^2}{1+\chi q(q-1)\big|\eta_q(B/\sigma + Z;\chi)\big|^{q-2}}\right]}_{I_1} + \underbrace{\mathbb{E}\left[\frac{BZ\partial_1\eta_q(B/\sigma + Z;\chi)}{\sigma\chi}\right]}_{I_2}$$
$$\underbrace{-\frac{1}{\sigma\chi}\mathbb{E}\left[\frac{B\eta_q(B/\sigma + Z;\chi)}{1+\chi q(q-1)\big|\eta_q(B/\sigma + Z;\chi)\big|^{q-2}}\right]}_{I_3}.$$

By a similar argument and using DCT, it is not hard to see that,

$$\lim_{\sigma\to\infty} \sigma^2\chi^{\frac{q}{q-1}} I_1 = \frac{\mathbb{E}B^2\mathbb{E}|Z|^{\frac{2-q}{q-1}}}{q^{\frac{1}{q-1}}(q-1)}, \tag{4.16}$$
$$\lim_{\sigma\to\infty} \sigma\chi^{\frac{q+1}{q-1}} I_3 = \frac{\mathbb{E}B\mathbb{E}\big(|Z|^{\frac{3-q}{q-1}}\operatorname{sgn}(Z)\big)}{q^{\frac{2}{q-1}}(q-1)} = 0.$$

Regarding the term I_2, by using Stein's lemma and Taylor expansion, we can obtain a sequel of equalities:

$$
\begin{aligned}
I_2 &= \frac{\mathbb{E}\big[B(Z^2-1)\eta_q\big(B/\sigma+Z;\chi\big)\big]}{\chi\sigma} \\
&= \frac{\mathbb{E}\big[B(Z^2-1)\big(\eta_q(Z;\chi)+\partial_1\eta_q(\gamma B/\sigma+Z;\chi)B/\sigma\big)\big]}{\chi\sigma} \\
&= \frac{\mathbb{E}\big[B^2(Z^2-1)\partial_1\eta_q\big(\gamma B/\sigma+Z;\chi\big)\big]}{\chi\sigma^2} \\
&= \frac{1}{\chi\sigma^2}\mathbb{E}\left[\frac{B^2(Z^2-1)}{1+\chi q(q-1)\big|\eta_q\big(\gamma B/\sigma+Z;\chi\big)\big|^{q-2}}\right],
\end{aligned}
$$

where the second step is simply due to Lemma 2.5.5 part (iv); $\gamma\in(0,1)$ is a random variable depending on B and Z. Again with a similar argument to verify the conditions of DCT we obtain

$$
\lim_{\sigma\to\infty}\chi^{\frac{q}{q-1}}\sigma^2 I_2 = \frac{(2-q)\mathbb{E}B^2\mathbb{E}|Z|^{\frac{2-q}{q-1}}}{q^{\frac{1}{q-1}}(q-1)^2}. \tag{4.17}
$$

Finally, based on (4.14) and collecting the results from (4.15), (4.16) and (4.17) enables us to have as $\sigma\to\infty$,

$$
\frac{\chi}{\sigma^{2(q-1)}} = \left[\frac{\chi^{\frac{q+1}{q-1}}(I_3-H_1)}{\sigma^2\chi^{\frac{q}{q-1}}(I_1+I_2)}\right]^{q-1} \to \left(\frac{q-1}{q^{\frac{1}{q-1}}}\frac{\mathbb{E}|Z|^{\frac{2}{q-1}}}{\mathbb{E}B^2\mathbb{E}|Z|^{\frac{2-q}{q-1}}}\right)^{q-1}.
$$

□

We now characterize the convergence rate of $R_q(\chi_q^*(\sigma),\sigma)$ in the lemma below.

Lemma 4.3.6. *Suppose* $1<q\le 2$ *and* G *has finite moments of all orders, then as* $\sigma\to\infty$*,*

$$
R_q(\chi_q^*(\sigma),\sigma) = \frac{\epsilon\mathbb{E}|G|^2}{\sigma^2} - \frac{\epsilon^2\big(\mathbb{E}|G|^2\mathbb{E}|Z|^{\frac{2-q}{q-1}}\big)^2}{(q-1)^2\mathbb{E}|Z|^{\frac{2}{q-1}}}\frac{1}{\sigma^4} + o(1/\sigma^4).
$$

Proof. It is straightforward to prove the result for $q=2$. From now on we only consider $1<q<2$. We write χ for $\chi_q^*(\sigma)$ in the rest of the proof to simplify the notation. First we have

$$
\begin{aligned}
R_q(\chi,\sigma) - \frac{\epsilon\mathbb{E}|G|^2}{\sigma^2} &= \mathbb{E}\eta_q^2(B/\sigma+Z;\chi) - 2\mathbb{E}[\eta_q(B/\sigma+Z;\chi)B/\sigma] \\
&= \mathbb{E}\eta_q^2(B/\sigma+Z;\chi) - 2\mathbb{E}[(\eta_q(Z;\chi)+\partial_1\eta_q(\gamma B/\sigma+Z;\chi)B/\sigma)B/\sigma] \\
&= \mathbb{E}\eta_q^2(B/\sigma+Z;\chi) - 2\mathbb{E}[\partial_1\eta_q(\gamma B/\sigma+Z;\chi)B^2/\sigma^2],
\end{aligned} \tag{4.18}
$$

where we have used Taylor expansion in the second step and $\gamma \in (0,1)$ is a random variable depending on B, Z. According to Lemma 2.5.5 part (i),

$$\begin{aligned}
\chi^{\frac{2}{q-1}} \eta_q^2(B/\sigma + Z;\chi) &= q^{\frac{2}{1-q}} (|B/\sigma + Z| - |\eta_q(B/\sigma+Z;\chi)|)^{\frac{2}{q-1}} \\
&\le q^{\frac{2}{1-q}} (|B| + |Z|)^{\frac{2}{q-1}}, \quad \text{for } \sigma \ge 1.
\end{aligned}$$

The upper bound is integrable since G has finite moments of all orders. Hence we can apply DCT to obtain

$$\lim_{\sigma\to\infty} \chi^{\frac{2}{q-1}} \mathbb{E}\eta_q^2(B/\sigma + Z;\chi) = q^{\frac{2}{1-q}} \mathbb{E}|Z|^{\frac{2}{q-1}}. \tag{4.19}$$

We can follow a similar argument to use DCT to have

$$\begin{aligned}
&\lim_{\sigma\to\infty} \chi^{\frac{2}{q-1}} \mathbb{E}[\partial_1 \eta_q(\gamma B/\sigma + Z;\chi) B^2/\sigma^2] \\
\overset{(a)}{=}\ &\lim_{\sigma\to\infty} \frac{\chi^{\frac{1}{q-1}}}{\sigma^2} \cdot \lim_{\sigma\to\infty} \mathbb{E}\left[\frac{B^2}{\chi^{-\frac{1}{q-1}} + q(q-1)|\chi^{\frac{1}{q-1}}\eta_q(\gamma B/\sigma + Z;\chi)|^{q-2}}\right] \\
\overset{(b)}{=}\ &\frac{q-1}{q^{\frac{1}{q-1}}} \frac{\mathbb{E}|Z|^{\frac{2}{q-1}}}{\mathbb{E}B^2 \mathbb{E}|Z|^{\frac{2-q}{q-1}}} \cdot \frac{\mathbb{E}B^2 \mathbb{E}|Z|^{\frac{2-q}{q-1}}}{q^{\frac{1}{q-1}}(q-1)} = q^{\frac{2}{1-q}} \mathbb{E}|Z|^{\frac{2}{q-1}},
\end{aligned} \tag{4.20}$$

where (a) holds due to Lemma 2.5.8 part (i); we have used Lemma 4.3.5 and DCT to obtain (b). Finally, we put the results (4.18), (4.19), (4.20) and Lemma 4.3.5 together to derive

$$\begin{aligned}
&\lim_{\sigma\to\infty} \sigma^4 (R_q(\chi,\sigma) - \epsilon \mathbb{E}|G|^2/\sigma^2) \\
=\ &\lim_{\sigma\to\infty} \frac{\sigma^4}{\chi^{\frac{2}{q-1}}} \cdot \Big[\lim_{\sigma\to\infty} \chi^{\frac{2}{q-1}} \mathbb{E}\eta_q^2(B/\sigma + Z;\chi) - \\
&\qquad 2\lim_{\sigma\to\infty} \chi^{\frac{2}{q-1}} \mathbb{E}(\partial_1\eta_q(\gamma B/\sigma + Z;\chi) B^2/\sigma^2)\Big] \\
=\ &\left(\frac{q-1}{q^{\frac{1}{q-1}}} \frac{\mathbb{E}|Z|^{\frac{2}{q-1}}}{\mathbb{E}B^2\mathbb{E}|Z|^{\frac{2-q}{q-1}}}\right)^{-2} \cdot (q^{\frac{2}{1-q}} \mathbb{E}|Z|^{\frac{2}{q-1}} - 2q^{\frac{2}{1-q}}\mathbb{E}|Z|^{\frac{2}{q-1}}) \\
=\ &-\frac{\epsilon^2 \big(\mathbb{E}|G|^2 \mathbb{E}|Z|^{\frac{2-q}{q-1}}\big)^2}{(q-1)^2 \mathbb{E}|Z|^{\frac{2}{q-1}}}.
\end{aligned}$$

This finishes the proof. □

4.3.3.2 Deriving the expansion of $\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w)$ for $q \in (1, 2]$

The way we derive the result (4.2) of Theorem 4.2.1 is similar to that in Chapter 4.3.2.3. We hence do not repeat all the details. The key step is applying Lemma 4.3.6 to obtain

$$
\begin{aligned}
& \lim_{\sigma_w \to \infty} \sigma_w^2 (\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) - \epsilon \mathbb{E}|G|^2) = \lim_{\bar{\sigma} \to \infty} \sigma_w^2 (\mathrm{AMSE}(\lambda_{*,q}, q, \sigma_w) - \epsilon \mathbb{E}|G|^2) \\
= & \lim_{\bar{\sigma} \to \infty} \frac{\sigma_w^2}{\bar{\sigma}^2} \cdot \lim_{\bar{\sigma} \to \infty} \bar{\sigma}^4 (R_q(\chi_q^*(\bar{\sigma}), \bar{\sigma}) - \epsilon \mathbb{E}|G|^2 / \bar{\sigma}^2) \\
= & -\epsilon^2 (\mathbb{E}|G|^2)^2 c_q.
\end{aligned}
$$

4.3.4 Proof of Theorem 4.2.1 for $q > 2$

We aim to prove the same results as presented in Lemmas 4.3.5 and 4.3.6. However, many of the limits we took when proving for the case $1 < q \le 2$ become invalid for $q > 2$ because DCT may not be applicable. Therefore, here we assume a slightly stronger condition that G has a sub-Gaussian tail and use a different reasoning to validate the results in Lemmas 4.3.5 and 4.3.6. Throughout this section, we use χ to denote $\chi_q^*(\sigma)$ for simplicity. First note that Lemma 4.3.2 holds for $q > 2$ as well. Hence we already know $\chi_q^*(\sigma) \to \infty$ as $\sigma \to \infty$. The following key lemma paves our way for the proof.

Lemma 4.3.7. *Suppose function $h : \mathbb{R}^2 \to \mathbb{R}$ satisfies $|h(x, y)| \le C(|x|^{m_1} + |y|^{m_2})$ for some $C > 0$ and $0 \le m_1, m_2 < \infty$. B has sub-Gaussian tail. Then the following result holds for any constants $v \ge 0, \gamma \in [0, 1]$ and $q > 2$,*

$$
\begin{aligned}
& \lim_{\sigma \to \infty} \chi^{\frac{v+1}{q-1}} \mathbb{E}\left[\frac{h(B, Z) |\eta_q(B/\sigma + Z; \chi)|^v}{1 + \chi q(q-1) |\eta_q(\gamma B/\sigma + Z; \chi)|^{q-2}} \right] \qquad (4.21) \\
= & \frac{q^{\frac{-v-1}{q-1}}}{q-1} \mathbb{E}[h(B, Z) |Z|^{\frac{v+2-q}{q-1}}], \quad \text{as } \sigma \to \infty.
\end{aligned}
$$

Moreover, there is a finite constant K such that for sufficiently large σ,

$$
\max_{0 \le \gamma \le 1} \chi^{\frac{v+1}{q-1}} \mathbb{E}\left[\frac{|h(B, Z)| |\eta_q(B/\sigma + Z; \chi)|^v}{1 + \chi q(q-1) |\eta_q(\gamma B/\sigma + Z; \chi)|^{q-2}} \right] \le K. \qquad (4.22)
$$

Proof. Define

$$A = \{|\eta_q(\gamma B/\sigma + Z; \chi)| \le \frac{1}{2}|\gamma B/\sigma + Z|\}.$$

We evaluate the expectation on the set A and its complement A^c respectively. Recall we use p_B to denote the distribution of B. By a change of variable we then have

$$\begin{aligned}
&\mathbb{E}\left[\frac{h(B,Z)|\eta_q(B/\sigma + Z;\chi)|^v}{1+\chi q(q-1)|\eta_q(\gamma B/\sigma + Z;\chi)|^{q-2}}\right] \\
= &\int \frac{h(x, y-\gamma x/\sigma)|\eta_q(y+(1-\gamma)x/\sigma;\chi)|^v}{1+\chi q(q-1)|\eta_q(y;\chi)|^{q-2}}\phi(y-\gamma x/\sigma)dydp_B(x).
\end{aligned}$$

We have on $\mathbb{I}_{\{|\eta_q(y;\chi)|\le \frac{1}{2}|y|\}}$ when σ is large enough,

$$\begin{aligned}
&\frac{\chi^{\frac{v+1}{q-1}}|h(x,y-\gamma x/\sigma)|\cdot|\eta_q(y+(1-\gamma)x/\sigma;\chi)|^v}{1+\chi q(q-1)|\eta_q(y;\chi)|^{q-2}}\phi(y-\gamma x/\sigma) \\
\overset{(a)}{\le}\ &\frac{|h(x,y-\gamma x/\sigma)|\cdot|\chi^{\frac{1}{q-1}}\eta_q(y+(1-\gamma)x/\sigma;\chi)|^v}{q(q-1)|\chi^{\frac{1}{q-1}}\eta_q(y;\chi)|^{q-2}}\phi(y-\gamma x/\sigma) \\
\overset{(b)}{=}\ &\frac{q^{\frac{q-2-v}{q-1}}|h(x,y-\gamma x/\sigma)|\cdot|y+(1-\gamma)x/\sigma|^{\frac{v}{q-1}}}{q(q-1)(|y|-|\eta_q(y;\chi)|)^{\frac{q-2}{q-1}}}\phi(y/\sqrt{2})e^{-\frac{1}{4}(y-\frac{2\gamma x}{\sigma})^2+\frac{\gamma^2x^2}{2\sigma^2}} \\
\overset{(c)}{\le}\ &\frac{2^{\frac{q-2}{q-1}}q^{\frac{q-2-v}{q-1}}|h(x,y-\gamma x/\sigma)|\cdot|y+(1-\gamma)x/\sigma|^{\frac{v}{q-1}}}{q(q-1)\,|y|^{\frac{w}{q-1}}}\phi(y/\sqrt{2})e^{\frac{\gamma^2x^2}{2\sigma^2}} \\
\overset{(d)}{\le}\ &\frac{2^{\frac{q-2}{q-1}}q^{\frac{q-2-v}{q-1}}(|x|^{m_1}+(|y|+|x|)^{m_2})\cdot(|y|+|x|)^{\frac{v}{q-1}}}{q(q-1)\,|y|^{\frac{q-2}{q-1}}}\phi(y/\sqrt{2})e^{c_0x^2}.
\end{aligned}$$

We have used Lemma 2.5.5 part (i) to obtain $(a)(b)$; (c) is due to the condition $|\eta_q(y;\chi)| \le \frac{1}{2}|y|$; and (d) holds because of the condition on the function $h(x,y)$. Notice that the numerator of the upper bound is essentially a polynomial in $|x|$ and $|y|$. Since B has sub-Gaussian tail, if we choose c_0 small enough (when σ is sufficiently large), the integrability with respect to x can be guaranteed. The integrability w.r.t. y is clear since $(2-q)/(q-1) > -1$. Thus we can apply DCT to obtain

$$\begin{aligned}
&\lim_{\sigma\to\infty}\chi^{\frac{v+1}{q-1}}\mathbb{E}\left[\frac{h(B,Z)|\eta_q(B/\sigma+Z;\chi)|^v}{1+\chi q(q-1)|\eta_q(\gamma B/\sigma+Z;\chi)|^{q-2}}\mathbb{I}_A\right] \\
= &\int \lim_{\sigma\to\infty}\frac{h(x,y-\gamma x/\sigma)|\chi^{\frac{1}{q-1}}\eta_q(y+(1-\gamma)x/\sigma;\chi)|^v}{\chi^{\frac{1}{1-q}}+q(q-1)|\chi^{\frac{1}{q-1}}\eta_q(y;\chi)|^{q-2}}\phi(y-\gamma x/\sigma)\mathbb{I}_{\{|\eta_q(y;\chi)|\le\frac{1}{2}|y|\}}dydp_B(x) \\
= &\int \frac{q^{\frac{-1-v}{q-1}}h(x,y)}{(q-1)|y|^{\frac{q-2-v}{q-1}}}\phi(y)dydp_B(x) = \frac{q^{\frac{-1-v}{q-1}}}{q-1}\mathbb{E}[h(B,Z)|Z|^{\frac{v+2-q}{q-1}}].
\end{aligned}$$

We now evaluate the expectation on the event A^c. Note that A^c implies

$$\begin{aligned}|\gamma B/\sigma + Z| &= \chi q|\eta_q(\gamma B/\sigma + Z;\chi)\Big|^{q-1} + |\eta_q(\gamma B/\sigma + Z;\chi)| \\ &> \frac{\chi q}{2^{q-1}}|\gamma B/\sigma + Z|^{q-1} + \frac{1}{2}|\gamma B/\sigma + Z| \\ &\Rightarrow |\gamma B/\sigma + Z| < 2(\chi q)^{\frac{1}{2-q}}.\end{aligned}$$

Hence we have the following bounds,

$$\begin{aligned}&\chi^{\frac{v+1}{q-1}}\mathbb{E}\left[\frac{|h(B,Z)|\cdot|\eta_q(B/\sigma+Z;\chi)|^v}{1+\chi q(q-1)|\eta_q(\gamma B/\sigma+Z;\chi)|^{q-2}}\mathbb{I}_{A^c}\right] \\ &\leq \chi^{\frac{v+1}{q-1}}\mathbb{E}(|h(B,Z)|\cdot|\eta_q(B/\sigma+Z;\chi)|^v\mathbb{I}_{A^c}) \\ &\leq \chi^{\frac{1}{q-1}}\int_{|y|<2(\chi q)^{\frac{1}{2-q}}}|h\big(x,y-\frac{\gamma x}{\sigma}\big)|\cdot|\chi^{\frac{1}{q-1}}\eta_q\big(y+\frac{1-\gamma}{\sigma}x;\chi\big)|^v\phi(y)e^{\frac{\gamma y x}{\sigma}}dydp_B(x) \\ &\overset{(e)}{\leq} q^{\frac{v}{1-q}}\chi^{\frac{1}{q-1}}\int_{|y|<2(\chi q)^{\frac{1}{2-q}}}(|x|^{m_1}+(|y|+|x|)^{m_2})\,(|y|+|x|)^{\frac{v}{q-1}}\phi(y)e^{\frac{2(\chi q)^{\frac{1}{2-q}}}{\sigma}x}dydp_B(x) \\ &\leq q^{\frac{v}{1-q}}\chi^{\frac{1}{q-1}}\int_{|y|<2(\chi q)^{\frac{1}{2-q}}}P(|x|,|y|)\phi(y)e^{\frac{2(\chi q)^{\frac{1}{2-q}}}{\sigma}x}dydp_B(x) \\ &\overset{(f)}{\leq} c_1\chi^{\frac{1}{q-1}}\chi^{\frac{1}{2-q}}\int\tilde{P}(|x|)e^x dp_B(x) \leq c_2\chi^{\frac{-1}{(q-1)(q-2)}} \to \infty \text{ as } \sigma\to\infty,\end{aligned}$$

where (e) is due to Lemma 2.5.5 part (i) and condition on $h(x,y)$; $P(\cdot,\cdot),\tilde{P}(\cdot)$ are two polynomials; the extra term $\chi^{\frac{1}{2-q}}$ in step (f) is derived from the condition $|y| < 2(\chi q)^{\frac{1}{2-q}}$. We thus have finished the proof of (4.21). Finally, note that the two upper bounds we derived do not depend on γ, hence (4.22) follows directly. □

We are now in position to prove Theorem 4.2.1 for $q > 2$. We will be proving the results of Lemmas 4.3.5 and 4.3.6 for $q > 2$. After that the exactly same arguments presented in Chapter 4.3.3.2 will close the proof. Since the basic idea of proving Lemmas 4.3.5 and 4.3.6 for $q > 2$ is the same as for the case $q \in (1, 2]$, we do not detail out the entire proof and instead highlight the differences. The major difference is that we apply Lemma 4.3.7 to make some of the limiting arguments valid in the case $q > 2$. Adopting the same notations in Chapter 4.3.3.1, we list the settings in the use of Lemma 4.3.7 below

- Lemma 4.3.5 I_1: set $h(x,y) = x^2, v = 0, \gamma = 1$.

- Lemma 4.3.5 I_3: set $h(x,y) = x\text{sgn}(\frac{x}{\sigma} + y), v = 1, \gamma = 1$. Note that the dependence of $h(x,y)$ on σ does not affect the result.

- Lemma 4.3.5 I_2: Notice we have

$$\begin{aligned}
\chi^{\frac{q}{q-1}}\sigma^2 I_2 =& \chi^{\frac{1}{q-1}}\sigma\mathbb{E}\left[B(Z^2-1)\Big(\eta_q(Z;\chi) + \frac{B}{\sigma}\int_0^1 \partial_1\eta_q(sB/\sigma + Z;\chi)ds\Big)\right] \\
=& \chi^{\frac{1}{q-1}}\int_0^1 \mathbb{E}\Big[B^2(Z^2-1)\partial_1\eta_q(sB/\sigma + Z;\chi)\Big]ds \\
=& \int_0^1 \chi^{\frac{1}{q-1}}\mathbb{E}\left[\frac{B^2(Z^2-1)}{1+\chi q(q-1)\big|\eta_q(sB/\sigma+Z;\chi)\big|^{q-2}}\right]ds.
\end{aligned}$$

We have switched the integral and expectation in the second step above due to the integrability. Set $h(x,y) = x^2(y^2-1), v = 0, \gamma = s$; then by the bound (4.22) in Lemma 4.3.7, we can bring the limit $\sigma \to \infty$ inside the above integral to obtain the result of $\lim_{\sigma\to\infty}\chi^{\frac{q}{q-1}}\sigma^2 I_2$.

- In Lemma 4.3.6, we need rebound the term $\mathbb{E}[\eta_q(B/\sigma + Z;\chi)B/\sigma]$ in (4.18).

$$\begin{aligned}
& \chi^{\frac{1}{q-1}}\sigma^2\mathbb{E}[\eta_q(B/\sigma + Z;\chi)B/\sigma] \\
=& \chi^{\frac{1}{q-1}}\sigma^2\mathbb{E}\left[\frac{B}{\sigma}\Big(\eta_q(Z;\chi) + \frac{B}{\sigma}\int_0^1 \partial_1\eta_q(sB/\sigma + Z;\chi)ds\Big)\right] \\
=& \int_0^1 \chi^{\frac{1}{q-1}}\mathbb{E}\left[\frac{B^2}{1+\chi q(q-1)\big|\eta_q(sB/\sigma+Z;\chi)\big|}\right]ds
\end{aligned}$$

We set $h(x,y) = x^2, v = 0, \gamma = s$. The rest arguments are similar to the previous one.

Chapter 5

Discussions

5.1 Linear asymptotics for non-convex problems

We remind the reader that our derivation of AMSE for LQLS with $q \in [0, 1)$ is based on the replica method from statistical physics. In particular, we have made the "replica symmetry" (RS) assumption for the method to work out the existing AMSE formula. A crucial follow-up procedure is to perform the "replica-symmetry-breaking" (RSB) scheme to check the validity of the RS assumption, and potentially improve over the RS result. Since the RSB calculations are far more complicated than RS calculations, we leave the work for future research.

Under our linear asymptotic ($n/p \to \delta$) framework, we should realize that the AMSE characterization of LQLS for $0 \leq q < 1$ (we have to resort to the non-rigorous replica method) is much harder than that for $q \geq 1$ (we have rigorous and general results). There seemingly exists a theoretical barrier between convex and non-convex bridge regression problems. It seems such a barrier exists under more general non-convex regression problems. As far as we know there has not been a rigorous derivation of AMSE for any non-convex regression problems under the linear asymptotics. Establishing a fully rigorous treatment is an important and challenging research direction.

5.2 Variable selection via bridge regression

The thesis is focused on characterizing the behavior of bridge regression for the goal of parameter estimation. Hence we have used AMSE to measure the performance. However, if the primary interest lies on variable selection, how would bridge regression perform? Will the preceding analyses be irrelevant in this case? To answer these questions, we first define a measure of variable selection performance. For a given estimator $\hat{\beta}$, we consider the false discovery proportion (FDP) and true positive proportion (TPP), defined as

$$\text{FDP} = \frac{\sum_{i=1}^{p} \mathbb{1}(\hat{\beta}_i \neq 0, \beta_i = 0)}{\sum_{i=1}^{p} \mathbb{1}(\hat{\beta}_i \neq 0)}, \quad \text{TPP} = \frac{\sum_{i=1}^{p} \mathbb{1}(\hat{\beta}_i \neq 0, \beta_i \neq 0)}{\sum_{i=1}^{p} \mathbb{1}(\beta_i \neq 0)}.$$

We have been able to derive the asymptotically exact limits of FDP and TPP for LASSO and sparse estimators obtained by thresholding $\hat{\beta}(\lambda, q)$ with $q \geq 1$. The variable selection performance of LASSO or thresholded LQLS can then be evaluated by plotting TPP (the limit) against FDP (the limit) at various tuning or threshold levels. Denote this ROC-type curve by TF curve, and recall the optimal tuning $\lambda_{*,q}$ for AMSE

$$\lambda_{*,q} = \arg\min_{\lambda} \lim_{n\to\infty} \frac{1}{p}\|\hat{\beta}(\lambda, q) - \beta\|_2^2.$$

We have proved that, the thresholded $\hat{\beta}(\lambda_{*,1}, 1)$ achieves a uniformly improved TF curve than the plain LASSO. We further generalize the result to LQLS for any $q \geq 1$ and conclude that thresholded version of $\hat{\beta}(\lambda_{*,q}, q)$ with smaller AMSE attains a uniformly better TF curve, i.e., having an improved variable selection performance. The implication of our results is two-fold. Firstly, variable selection does not have to be carried out by sparsity inducing regularization in a single step. Rather, thresholding a regularized estimator may lead to superior performance, even if the regularized estimator is not sparse. For example, in the low signal-noise ratio case, thresholded ridge regression can obtain a better TF curve than LASSO. Secondly, the goal of variable selection is tightly aligned with parameter estimation, at least for the family of LQLS. Hence, our analysis of LQLS for parameter estimation carries over to the variable selection paradigm. Refer to Wang et al. (2017) for all the details.

5.3 Prediction in bridge regression

By far we have demonstrated that for a given $q \in [1, \infty)$, we would like to use $\hat{\beta}(\lambda_{*,q}, q)$ for parameter estimation, and the thresholded version for variable selection. If we move the focus to prediction, does $\lambda = \lambda_{*,q}$ remain the optimal tuning for $\hat{\beta}(\lambda, q)$? Denote the optimal tuning of $\hat{\beta}(\lambda, q)$ for prediction by

$$\lambda_{**,q} = \arg\min_{\lambda} \lim_{n \to \infty} \frac{1}{n} \|X\hat{\beta}(\lambda, q) - X\beta\|_2^2.$$

We have proved that $\lambda_{*,q} > \lambda_{**,q}$ for LASSO, $\lambda_{*,q} = \lambda_{**,q}$ for ridge regression, and there is no definite conclusion for any other $q \neq 1, 2$. Our findings not only single out the featured properties of LASSO and ridge regression in the LQLS family, but have a valuable implication for tuning parameter selection under the Stein's Unbiased Risk Estimate (SURE) (Stein, 1981) framework. Since SURE is an unbiased estimator for the expected in-sample prediction error $\mathbb{E}\|X\hat{\beta}(\lambda, q) - X\beta\|_2^2/n$, in the large sample regime our result indicates that SURE can not identify the optimal tuning of LASSO for parameter estimation, but it does for ridge regression. We further show that for LASSO, as ϵ, the sparsity level of the true coefficient β, goes to zero, $\lambda_{*,q} - \lambda_{**,q} \to 0$. The result suggests that in the extremely sparse scenarios, SURE can be safely used to select the optimal λ of LASSO for parameter estimation. We leave a full treatment regarding the prediction performance of bridge regression as a near future work.

Bibliography

Amelunxen, D., Lotz, M., McCoy, M. B., and Tropp, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, page iau005.

Bai, Z. and Yin, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The annals of Probability*, pages 1275–1294.

Bayati, M. and Montanari, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785.

Bayati, M. and Montanari, A. (2012). The lasso risk for gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017.

Bickel, P., Ritov, Y., and Tsybakov, A. (2009). Simultaneous analysis of lasso and dantzig selector. *Annals Stat.*, pages 1705–1732.

Bogdan, M., van den Berg, E., Sabatti, C., Su, W., and Candès, E. (2015). Slope:adaptive variable selection via convex optimization. *Annals App. Stat.*, 9(3):1103.

Bradic, J. and Chen, J. (2015). Robustness in sparse linear models: relative efficiency based on robust approximate message passing. *arXiv preprint arXiv:1507.08726*.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Bunea, F., Tsybakov, A., and Wegkamp, M. (2007). Sparsity oracle inequalities for the lasso. *Electronic Journal of Statistics*, 1:169–194.

Candès, E. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592.

Candes, E. J. and Plan, Y. (2011). A probabilistic and ripless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254.

Candes, E. J. and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425.

Chartrand, R. and Staneva, V. (2008). Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(3):035020.

Coolen, A. C. (2005). *The Mathematical Theory of Minority Games: Statistical Mechanics of Interacting Agents (Oxford Finance Series)*. Oxford University Press, Inc.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106.

Donoho, D., Gavish, M., and Montanari, A. (2013). The phase transition of matrix recovery from gaussian measurements matches the minimax mse of matrix denoising. *Proceedings of the National Academy of Sciences*, 110(21):8405–8410.

Donoho, D. and Montanari, A. (2013). High dimensional robust m-estimation: Asymptotic variance via approximate message passing. *Prob. Theo. Rel. Fields*, pages 1–35.

Donoho, D. and Montanari, A. (2015). Variance breakdown of huber (m)-estimators: $n/p \to m$.

Donoho, D. L. (2004). For most underdetermined systems of linear equations, the minimal i1-norm near-solution approximates the sparsest near-solution. *Comm. Pure and Appl. Math.*

Donoho, D. L. (2006a). Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306.

Donoho, D. L. (2006b). High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete & Computational Geometry*, 35(4):617–652.

Donoho, D. L., Maleki, A., and Montanari, A. (2009). Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919.

Donoho, D. L., Maleki, A., and Montanari, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Transactions on Information Theory*, 57(10):6920–6941.

Donoho, D. L. and Tanner, J. (2005a). Neighborliness of randomly projected simplices in high dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9452–9457.

Donoho, D. L. and Tanner, J. (2005b). Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9446–9451.

El Karoui, N., Bean, D., Bickel, P. J., Lim, C., and Yu, B. (2013). On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Frank, L. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135.

Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics*, 7(3):397–416.

Guo, D. and Verdú, S. (2005). Randomly spread cdma: Asymptotics via statistical physics. *IEEE Transactions on Information Theory*, 51(6):1983–2010.

Hoerl, A. and Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Huang, J., Horowitz, J., and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals Stat.*, pages 587–613.

Kabashima, Y., Wadayama, T., and Tanaka, T. (2009). A typical reconstruction limit for compressed sensing based on lp-norm minimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(09):L09003.

Karoui, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*.

Kashin, B. S. (1977). Diameters of some finite-dimensional sets and classes of smooth functions. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 41(2):334–351.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of statistics*, pages 1356–1378.

Koltchinskii, V. (2009). Sparsity in penalized empirical risk minimization. In *Annales de l'IHP Probabilités et statistiques*, volume 45, pages 7–57.

Krzakala, F., Mézard, M., Sausset, F., Sun, Y., and Zdeborová, L. (2012). Statistical-physics-based reconstruction in compressed sensing. *Physical Review X*, 2(2):021005.

Maleki, A. (2010). *Approximate message passing algorithms for compressed sensing*. Ph.D. thesis, Stanford University.

Massoulié, L. (2014). Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM.

Mézard, M., Parisi, G., and Virasoro, M. (1987). *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Co Inc.

Mossel, E., Neeman, J., and Sly, A. (2013). A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*.

Mossel, E., Neeman, J., and Sly, A. (2015). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461.

Mousavi, A., Maleki, A., and Baraniuk, R. G. (2017). Consistent parameter estimation for lasso and approximate message passing. *Annals of Statistics*. To appear.

Oymak, S. and Hassibi, B. (2016). Sharp mse bounds for proximal denoising. *Foundations of Computational Mathematics*, 16(4):965–1029.

Oymak, S., Thrampoulidis, C., and Hassibi, B. (2013). The squared-error of generalized lasso: A precise analysis. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 1002–1009. IEEE.

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239.

Rangan, S., Fletcher, A. K., and Goyal, V. K. (2012). Asymptotic analysis of map estimation via the replica method and applications to compressed sensing. *IEEE TRANSACTIONS ON INFORMATION THEORY*, 58(3).

Raskutti, G., Wainwright, M., and Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over-balls. *IEEE Trans. Info. Theory*, 57(10).

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151.

STOJNIC, M. (2009). Various thresholds for ℓ_1-optimization in compressed sensing.

Stojnic, M. (2013). Lifting ℓ_q-optimization thresholds. *arXiv:1306.3976*.

Su, W., Bogdan, M., and Candes, E. (2015). False discoveries occur early on the lasso path. *arXiv preprint arXiv:1511.01957*.

Tanaka, T. (2002). A statistical-mechanics approach to large-system analysis of cdma multiuser detectors. *IEEE Transactions on Information theory*, 48(11):2888–2910.

Thrampoulidis, C., Abbasi, E., and Hassibi, B. (2016). Precise error analysis of regularized m-estimators in high-dimensions. *arXiv preprint arXiv:1601.06233*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, pages 267–288.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wang, M., Xu, W., and Tang, A. (2011). On the performance of sparse recovery via ℓ_p-minimization ($0 \leq p \leq 1$). *IEEE Transactions on Information Theory*, 57(11):7255–7278.

Wang, S., Weng, H., and Maleki, A. (2017). Which bridge estimator is optimal for variable selection? *arXiv preprint arXiv:1705.08617*.

Zheng, L., Maleki, A., Weng, H., Wang, X., and Long, T. (2017). Does ℓ_p-minimization outperform ℓ_1-minimization? *IEEE TRANSACTIONS ON INFORMATION THEORY*.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.