Masters Research Report

Asymptotic Error and Regularization Path in High-Dimensional LDA:
A Random Matrix Approach

Han Jun Ko

December 1, 2025

Abstract

We study regularized discriminant analysis (rLDA) in the high-dimension, low-sample-size (HDLSS) regime through the random-matrix limit

$$\mathrm{Err}(\hat{w}_\lambda) \;\to\; \Phi\big(-\Theta(\lambda)\big).$$

Our contributions are threefold. First, under $\Sigma = I$ we derive a piecewise closed form for the MDP limit $\Theta(0)$ across $\gamma := p/n < 1$, $\gamma = 1$, and $\gamma > 1$, yielding a theoretical error curve $\Phi(-\Theta(0))$ that peaks at $\gamma = 1$ and reconciles classical HDLSS observations. Second, we verify via Szegő arguments that Toeplitz covariance families (including AR(1) and Matérn) satisfy the spectral assumptions required by the limiting risk, thereby justifying their use in our computations. Third, we develop a numerically stable approximation pipeline for general Toeplitz Σ that evaluates $\Theta(\lambda)$ using almost-sure limits of Stieltjes functionals based on empirical spectral sums, and we map the optimizer $\lambda_{\mathrm{opt}}(\gamma, \rho)$. Empirically, λ_{opt} increases with the aspect ratio γ and decreases with the correlation strength ρ; moreover, $\Theta(\lambda_{\mathrm{opt}})$ is close to $\Theta(0)$ at small γ and close to $\Theta(\infty)$ at large γ, quantifying the transition between the MDP ($\lambda = 0$) and MD ($\lambda = \infty$) ends. Under Matérn covariance the gap between $\Phi(-\Theta(\lambda_{\mathrm{opt}}))$ and $\Phi(-\Theta(\infty))$ is uniformly smaller than under AR(1), indicating stronger effective smoothing of high-frequency directions. These patterns provide a simple, interpretable guide for choosing λ as a function of (γ, ρ) before resorting to cross-validation.

Keywords: high-dimensional classification; regularized LDA; random matrix theory; Stieltjes transform; Toeplitz covariance;

1 Introduction

High-dimensional classification often relies on regularization to mitigate singularity and overfitting. Classical choices such as ridge and ℓ_1-type penalties remain effective in practice, yet a number of phenomena unique to the high-dimension, low-sample-size (HDLSS) regime demand problem-specific understanding. In binary discrimination, Ahn and Marron (2010) observed the striking "data piling" behavior, and Lee et al. (2013) formalized the *maximal data piling* (MDP) direction that maximizes the separation between class means within the piling subspace. Let $X \in \mathbb{R}^{n\times p}$ be the training matrix, $Y \in \mathbb{R}^n$ the labels with entries in $\{\pm 1\}$, and for the balanced case let $X_{(1)}$ and $X_{(-1)}$ denote class-wise blocks. Writing

$$C := \begin{bmatrix} X_{(1)}^\top & X_{(-1)}^\top \end{bmatrix} \in \mathbb{R}^{p\times n}, \quad P := CC^+, \quad w := (\text{mean difference vector}), \quad v_{MD} := \frac{w}{\|w\|},$$

the MDP direction is

$$v_{MDP} \;\propto\; (I - P)\, w, \tag{1}$$

while the ridge-regularized LDA (rLDA) direction,

$$v(\lambda) \;:=\; (CC^\top + \lambda I)^{-1} w, \qquad \lambda > 0, \tag{2}$$

provides a convex continuum that links the two extremes: $v(0)$ recovers MDP and $v(\infty)$ recovers the mean-difference rule v_{MD} Lee et al. (2013); Dobriban and Wager (2018). Empirically, Ahn and Marron (2010) showed that the MDP classification error peaks around the aspect ratio $\gamma := p/n \approx 1$ and improves as within-feature correlation increases.

A general asymptotic framework for ridge-type methods was developed by Dobriban and Wager (2018), who expressed the limiting misclassification risk via Stieltjes transforms of spectral measures from random matrix theory (RMT) Bai and Silverstein (2010); Marchenko and Pastur (1967). Their formula makes it, in principle, possible to chart how the error of $v(\lambda)$ varies with (γ, λ) under structural assumptions on the covariance Σ. In practice, however, (i) closed-form expressions are scarce even for simple Σ, and (ii) a systematic, quantitative map of the *optimal* regularization λ_{opt} across *Toeplitz-type* covariances (e.g., AR(1), Matérn) as functions of (γ, ρ) has been largely implicit.

Research gap. Despite a well-developed general theory, there remains a gap between (a) the abstract limiting error formula and (b) concrete, closed-form or computationally stable guidance for λ selection in structured covariances. Moreover, the HDLSS literature documents empirical transitions between the MDP end ($\lambda = 0$) and the MD end ($\lambda = \infty$), but a concise, theory-backed account of when $v(\lambda)$ should behave "closer" to one end or the other—*as a function of* (γ, ρ)—is still fragmented across simulations and special cases.

Contributions. This paper addresses the above gap along three axes:

1. **Closed form for the MDP limit under $\Sigma = I$.** We derive a piecewise closed-form expression for the limiting margin $\Theta(0)$ of $v(\lambda)$ as $\lambda \to 0$ (i.e., the MDP limit) across the regimes $\gamma < 1$, $\gamma = 1$, and $\gamma > 1$. This yields a clean theoretical curve for the MDP error $\Phi(-\Theta(0))$ that reproduces and explains the "peak near $\gamma = 1$" observed in Ahn and Marron (2010). The derivation uses explicit Stieltjes and companion transforms for the Marchenko–Pastur law and is documented in the Appendix.

2. **Toeplitz covariances meet the assumptions of the RMT limit.** We verify via Szegő-type arguments that Toeplitz covariance families (including AR(1) and Matérn) satisfy the spectral assumptions required by Dobriban and Wager (2018). This justifies, in a self-contained way, applying their limit formula to these structured and practically important models.

3. **A stable procedure to compute $\Theta(\lambda)$ and map $\lambda_{\text{opt}}(\gamma, \rho)$.** Leveraging almost-sure convergence of Stieltjes functionals, we provide a numerically stable approximation pipeline for $\Theta(\lambda)$ and its maximizer λ_{opt}. We then chart the *quantitative* patterns: (i) λ_{opt} generally increases with γ but decreases with the correlation parameter ρ in AR(1) and Matérn; (ii) $\Theta(\lambda_{\text{opt}})$ is close to $\Theta(0)$ at small γ and approaches $\Theta(\infty)$ as γ grows, clarifying the empirical transition between the MDP and MD ends.

Main idea in one line. We start from the general $\text{Err}(v(\lambda)) \to \Phi(-\Theta(\lambda))$ limit Dobriban and Wager (2018) and make it *explicit and computable* at the two ends (MDP/MD) and *predictive* in between for Toeplitz covariances, thereby turning the abstract RMT formula into concrete guidance for λ selection as (γ, ρ) vary.

Organization. Section 2 reviews MDP/MD, key RMT tools, and the main assumptions. Section 3 gives the closed-form MDP limit under $\Sigma = I$ and reconciles it with the HDLSS patterns in Ahn and Marron (2010). Section 4 develops the approximation scheme for $\Theta(\lambda)$, verifies Toeplitz families (AR(1), Matérn), and presents the resulting maps of $\lambda_{\text{opt}}(\gamma, \rho)$ together with the transition behavior between $\lambda = 0$ and $\lambda = \infty$. The Appendix (Section 6) contains derivations and auxiliary lemmas.

2 Basic Setup

2.1 rLDA vs MDP vs MD

Let us see the relationship between rLDA, MDP, and mean difference. The rLDA direction vector is

$$v(\lambda) \propto (CC^T + \lambda I)^{-1} w \tag{3}$$

for $\lambda > 0$. For MDP,

$$v(\lambda) \propto \{I - I + \lambda(CC^T + \lambda I)^{-1}\} w \tag{4}$$

$$= \{I - (CC^T + \lambda I - \lambda I)(CC^T + \lambda I)^{-1}\} w \tag{5}$$

$$= \{I - CC^T(CC^T + \lambda I)^{-1}\}w \tag{6}$$

Now, consider the singular value decomposition

$$C = USV^T. \tag{7}$$

where $U \in \mathbb{R}^{p\times p}$ is an orthogonal matrix, $S \in \mathbb{R}^{p\times n}$ is a diagonal matrix with non-negative real numbers on the diagonal, and $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix. Then,

$$CC^T = USV^TVS^TU^T = USS^TU^T \tag{8}$$
$$CC^T + \lambda I = U(SS^T + \lambda I)U^T \tag{9}$$
$$(CC^T + \lambda I)^{-1} = U(SS^T + \lambda I)^{-1}U^T \tag{10}$$
$$C^T(CC^T + \lambda I)^{-1} = VS^T(SS^T + \lambda I)^{-1}U^T. \tag{11}$$

When $n \geq p$, the matrix SS^T becomes a $p \times p$ diagonal matrix whose jth diagonal is a square of jth singular value of C. When $n < p$, the matrix SS^T is a $p \times p$ diagonal matrix, whose jth diagonal is a square of jth singular value of C for $1 \leq j \leq n$, and 0 for $n + 1 \leq j \leq p$. Letting σ_j^2 be the j diagonal of SS^T,

$$(SS^T + \lambda I)^{-1} = \text{Diag}\left\{\frac{1}{\sigma_j^2 + \lambda}\right\}. \tag{12}$$

This makes the j disgonal of $S^T(SS^T + \lambda I)^{-1}$ be $\frac{\sigma_j}{\sigma_j^2+\lambda}$, so $\lim_{\lambda\to 0} S^T(SS^T + \lambda I)^{-1} = S^+$. Thus,

$$\lim_{\lambda\to 0} C^T(CC^T + \lambda I)^{-1} = VS^+U^T = C^+. \tag{13}$$

This leads to

$$\lim_{\lambda\to 0} \lambda(CC^T + \lambda I)^{-1}w = \lim_{\lambda\to 0}\{\ I - CC^T(CC^T + \lambda I)^{-1}\}w = (I - CC^+)w \propto v_{MDP}. \tag{14}$$

When it comes to mean difference,

$$v(\lambda) \propto \lambda(CC^T + \lambda I)^{-1}w \tag{15}$$
$$= \lambda(USS^TU^T + \lambda I)^{-1}w \tag{16}$$
$$= U\text{Diag}\left\{\frac{\lambda}{\sigma_j^2 + \lambda}\right\}U^Tw, \tag{17}$$

so,

$$\lim_{\lambda\to\infty} U\text{Diag}\left\{\frac{\lambda}{\sigma_j^2 + \lambda}\right\}U^Tw = UU^Tw = w \tag{18}$$

and $\lim_{\lambda\to\infty} v(\lambda)$ becomes MD.

2.2 Random Matrix Theory

This paper leverages several concepts of random matrix theory, which is widely used in high-dimensional asymptotics. The results are typically written using "empitical spectral distribution" (or spectral measure in random matrix), which, for a symmetric matrix $A \in \mathbb{R}^{p\times p}$, is a cumulative distribution function of A's eigenvalues:

$$F_A(s) := \frac{1}{p}\sum_{j=1}^{p} \mathbb{I}(\lambda_j(A) \leq s).$$

Using this spectral distribution, we have the "Marchenko Pastur" distribution.

Definition 1. *(Marchenko Pastur distribution, Marchenko and Pastur (1967)) Suppose X is $n \times p$ matrix and $F_{\hat{\Sigma}}$ is an empirical spectral distribution of $X^TX/n := \hat{\Sigma}$. Then $F_{\hat{\Sigma}}$ converges almost surely to some F as $n, p \to \infty$, where $p/n \to \gamma$ and F is the CDF of the following pdf f:*

$$f(s) = \frac{1}{2\pi\gamma s}\sqrt{(b - s)(s - a)},$$

$$a = (1 - \sqrt{\gamma})^2, \quad b = (1 + \sqrt{\gamma})^2.$$

The Stieltjes transform of MP distribution can be written as

$$\int \frac{1}{s - z} dF(s) := m(z).$$

The companion Stieltjes transform $v(z)$ of the Marchenko Pastur distribution is defined as follows J.W.Silverstein (1995):

$$v(z) + \frac{1}{z} = \gamma\left(m(z) + \frac{1}{z}\right). \tag{19}$$

The Stieltjes transform of Marchenko Pastur distribution and its companion Stieltjes transform under identity covariance $\Sigma = I$ are as follows (Bai and Silverstein (2010)) :

$$m(z) = \frac{1 - \gamma - z - \sqrt{(1 - \gamma - z)^2 - 4\gamma z}}{2\gamma z} \tag{20}$$

$$v(z) = \frac{\gamma - 1 - z - \sqrt{(1 - \gamma - z)^2 - 4\gamma z}}{2z}. \tag{21}$$

2.3 Data Generation

As mentioned, this covers the binary classification problem using training data $X \in \mathbb{R}^{n \times p}$ with population covariance $\Sigma \in \mathbb{R}^{p \times p}$ and $Y \in \mathbb{R}^n$. For simplicity, this paper considers the case where the number of class 1 and -1 is identical. Since we make use of the classification error formula in Dobriban and Wager (2018), we use the same assumption in the paper to generate data, especially based on 1st assumption.

Assumption A. *The following conditions hold.*

1. *The data X is generated as $X = Z\Sigma^{1/2}$ for an $n \times p$ matrix Z with i.i.d entries satisfying $\mathbb{E}[Z_{ij}] = 0$ and $Var[Z_{ij}] = 1$, and a deterministic $p \times p$ positive definite matrix Σ.*
2. *$n, p \to \infty$ so that $p/n \to \gamma > 0$*
3. *The empirical spectral distribution F_Σ converges to a limit probability distribution H supported on $[0, \infty)$.*

2.4 Why AR(1) and Matérn?

Toeplitz-type covariances are ubiquitous in practice and form a mathematically tractable class under which the spectral assumptions in Assumption A are satisfied via Szegő-type results. We focus on two representative families:

- **AR(1)**: $(\Sigma_\rho)_{ij} = \rho^{|i-j|}$ with $\rho \in (0, 1)$ captures short-range temporal correlation with a single interpretable parameter ρ. Its Toeplitz structure yields a closed-form spectral density, enabling stable evaluation of Stieltjes functionals.
- **Matérn** ($\nu > 0$, range $\rho > 0$): $C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\, \frac{d}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\, \frac{d}{\rho}\right)$ with $d = |i - j|$. It generalizes exponential ($\nu = \frac{1}{2}$) and approximates Gaussian ($\nu \to \infty$), covering a wide spectrum of smoothness levels in spatial statistics.

These two choices jointly probe how *range (AR)* and *smoothness (Matérn)* impact the regularization path $\lambda \mapsto \Theta(\lambda)$ and the optimizer λ_{opt}.

2.5 Computing $\Theta(\lambda)$ and λ_{opt} (Stable A.S.-Limit Approximation)

Let $\widehat{\Sigma} := X^T X/n$ with eigenvalues $\{\lambda_j\}_{j=1}^p$. For $z = -\lambda < 0$, approximate Stieltjes quantities by

$$m(z) \approx \frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j - z}, \qquad v(z) \approx \gamma \left(\frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j - z} + \frac{1}{z}\right) - \frac{1}{z},$$

and derivatives by

$$m'(z) \approx \frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2}, \qquad v'(z) \approx \gamma\left(\frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2} - \frac{1}{z^2}\right) + \frac{1}{z^2}.$$

Plug these into the standard decomposition

$$\tau(\lambda) = \lambda\, m(-\lambda)\, v(-\lambda), \quad \eta(\lambda) = \frac{v(-\lambda) - \lambda v'(-\lambda)}{\gamma}, \quad \xi(\lambda) = \frac{v'(-\lambda)}{v(-\lambda)^2} - 1,$$

and then compute

$$\Theta(\lambda) \;=\; \frac{\alpha^2\,\tau(\lambda)}{\sqrt{\alpha^2\,\eta(\lambda) + \xi(\lambda)}}, \qquad \mathrm{Err}(v(\lambda)) \;\approx\; \Phi(-\Theta(\lambda))\,.$$

3 Classification Error of MDP

Throughout the article Dobriban and Wager (2018), it states that the classification error of rLDA converges as follows:

$$\mathrm{Err}(\hat{w}_\lambda) \xrightarrow{a.s.} \Phi(-\Theta(\lambda)), \quad \text{where} \tag{22}$$

$$\Theta(\lambda) = \frac{\alpha^2\tau(\lambda)}{\sqrt{\alpha^2\eta(\lambda) + \xi(\lambda)}}, \tag{23}$$

$$\tau(\lambda) = \lambda m v, \quad \eta(\lambda) = \frac{v - \lambda v'}{\gamma}, \quad \xi(\lambda) = \frac{v'}{v^2} - 1. \tag{24}$$

The m and v have closed form under identity covariance, as written in (20) and (21). Not only that, the identity covariance satisfies the conditions of the classification error formula, especially regarding the eigenvalue condition of covariance matrix that requires them to be between some positive values. So, this section shows how the classification error of MDP, where $\lambda \to 0$, looks like under identity covariance and reconcile it with the HDLSS patterns reported in Ahn and Marron (2010).

3.1 Derivation under Identity Covariance

Recall the $\Theta(\lambda)$ expression

$$\Theta(\lambda) = \frac{\alpha^2\tau(\lambda)}{\sqrt{\alpha^2\eta(\lambda) + \xi(\lambda)}}, \tag{25}$$

Then, we can obtain $\lim_{\lambda\to 0}\Theta(\lambda) := \Theta(0)$ under identity covariance in a closed form as follows.

Corollary 1.

$$\Theta(0) = \begin{cases} \dfrac{\alpha^2}{\sqrt{\alpha^2\frac{1}{1-\gamma} + \frac{\gamma}{1-\gamma}}} & ,0 < \gamma < 1 \\ 0 & ,\gamma = 1 \\ \dfrac{\alpha^2}{\sqrt{\alpha^2\frac{\gamma}{\gamma-1} + \frac{\gamma^2}{\gamma-1}}} & ,\gamma > 1. \end{cases}$$

The **Appendix** in Section 6 provides how it is derived.

3.2 Result

The following Figure 1 shows the graphs of $\Phi(-\Theta(0))$ when $\Sigma = I$ for different values of $\alpha = 1, 2, 3$. Commonly, all the graphs increases rapidly in $\gamma \in (0, 1)$ and reach the maximum $0.5 = \Phi(0)$ at $\gamma = 1$. After they attain their local minimum, the value increases gradually to maximum 0.5. As the α increases, the classification error decreases at the same γ.

This result complies with that of Ahn and Marron (2010) that also explains the classification error vs dimension (instead of γ) of maximal data piling under compound symmetry covariance. Looking the case $\rho = 0$ that corresponds to identity matrix, you can see that the ascend-descend-reascend pattern also appears in Ahn and Marron (2010), which is analogous to our result. We now situate these closed-form curves within the prior HDLSS literature to clarify what is new in our treatment.

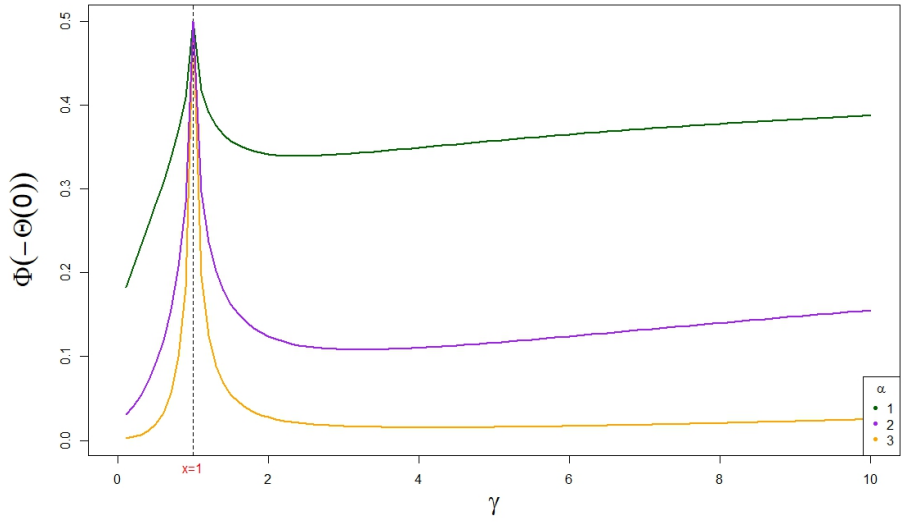

Figure 1

3.3 Comparison with Prior Work

Ahn & Marron (2010). Under compound symmetry (including the $\rho = 0$ identity case), they reported via *simulation* that MDP's error *rises–falls–rises* with a peak near $\gamma \approx 1$. Here, under $\Sigma = I$ we derive a *closed-form* $\Theta(0)$ that (i) *reconstructs* the same pattern theoretically and (ii) *pinpoints* the peak magnitude (0.5) and location ($\gamma = 1$).

4 Minimum of Risk

In this section, we'll delve into the λ at which the classification error becomes minimum, that is, the value of $\Theta(\lambda)$ becomes maximum. We'll denote this as λ_{opt}.

4.1 Approximation of $\Theta(\lambda)$

Except for special cases (i.e. Identity covariance), the Stieltjes transform of Marchenko Pastur distribution $m(z)$ and its companion Stieltjes transform $v(z)$ do not have explicit closed form regarding $z = -\lambda$. So, we utilize almost sure convergence property of $m(z)$, which is also presented in the introduction of Dobriban and Wager (2018), to approximate the $\Theta(\lambda)$.

Recall the definition of Marchenko Pastur distribution in **Definition 1**. Using the definition of the empirical spectral distribution, we have

$$\int_{s=0}^{\infty} \frac{1}{s-z} dF_{\hat{\Sigma}}(s) = \frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j - z} \overset{a.s.}{\to} m(z), \tag{26}$$

where λ_j is the jth eigenvalue of $\hat{\Sigma}$. This convergence result can also be applied to its companion Stieltjes transform (19) so that

$$\gamma \left(\frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j - z} + \frac{1}{z} \right) - \frac{1}{z} \overset{a.s.}{\to} v(z). \tag{27}$$

The almost sure convergence also applies to their derivatives.

$$\frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2} \overset{a.s.}{\to} m'(z), \quad \gamma\left(\frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2} - \frac{1}{z^2}\right) + \frac{1}{z^2} \overset{a.s.}{\to} v'(z). \tag{28}$$

We use this convergence results to approximate the value of $\Theta(\lambda)$ in (22).

4.2 Toeplitz justification via Szegő Theorem

The covariance matrices should satisfy the conditions in **Assumption** A of Dobriban and Wager (2018). Especially, regarding the 3rd condition, all matrices of **Toeplitz** form satisfies the given condition. This is guaranteed by so-called **Szegő's theorem**, which is as follows:

Theorem 1. *(Szegő's theorem) Let w be a Fourier series with Fourier coefficients c_k*

$$w(\theta) = \sum_{k=-\infty}^{\infty} c_k e^{ik\theta}, \quad \theta \in [0, 2\pi]$$
$$c_k = \frac{1}{2\pi}\int_0^{2\pi} w(\theta)e^{-ik\theta}d\theta,$$

$\Sigma(w) := \Sigma$ is a $p \times p$ Toeplitz matrix, and $\lambda_j, j = 1, \ldots, p$ is jth eigenvalue of Σ. Then, for any function G that is continuous on the range of w,

$$\lim_{p\to\infty}\frac{1}{p}\sum_{j=1}^{p} G(\lambda_j) = \frac{1}{2\pi}\int_0^{2\pi} G(w(\theta))d\theta. \tag{29}$$

Based on the Toeplitz matrix theory, the spectral density f_Σ can be expressed as a Fourier series as follows:

$$f_\Sigma(\theta) = \sum_{d=-\infty}^{\infty} C(d)e^{-id\theta}$$

where $C(|i-j|) := \Sigma_{i,j}$. Then, applying $G(\cdot) := \mathbb{I}(\cdot \le x)$ to the equation (29) leads to

$$\frac{1}{p}\sum_{j=1}^{p}\mathbb{I}(\lambda_j \le x) = F_\Sigma(x) \to \frac{1}{2\pi}\int_0^{2\pi}\mathbb{I}(f_\Sigma(\theta) \le x)d\theta := H(x). \tag{30}$$

Defining the H as in the equation (30), this concludes that the Toeplitz form satisfies the **Assumption** A of Dobriban and Wager (2018).

The **Toeplitz** form includes identity matrix, AR(1), Matern covariance, and so on. From next section, experiments will be performed for the two cases AR(1) and Matern covariance, which are included in Toeplitz matrix form.

Computation recipe (one setting (γ, ρ))

1. Generate balanced labels and $X = Z\Sigma^{1/2}$ with the target Toeplitz Σ (AR(1) or Matérn), repeat 20 times.
2. For each replicate, compute eigenvalues of the pooled within–class covariance and evaluate $m(-\lambda), v(-\lambda)$ and derivatives via (26)–(28).
3. Form $\Theta(\lambda)$ from (22) on a log grid; refine the maximizer to obtain λ_{opt}.
4. Report the replicate–averaged $\Theta(\lambda_{\text{opt}})$, $\Theta(\lambda_0)$, and $\Theta(\lambda_\infty)$; plot error as $\Phi(-\Theta)$ (and optionally $\log\Phi(-\Theta)$ for visibility).

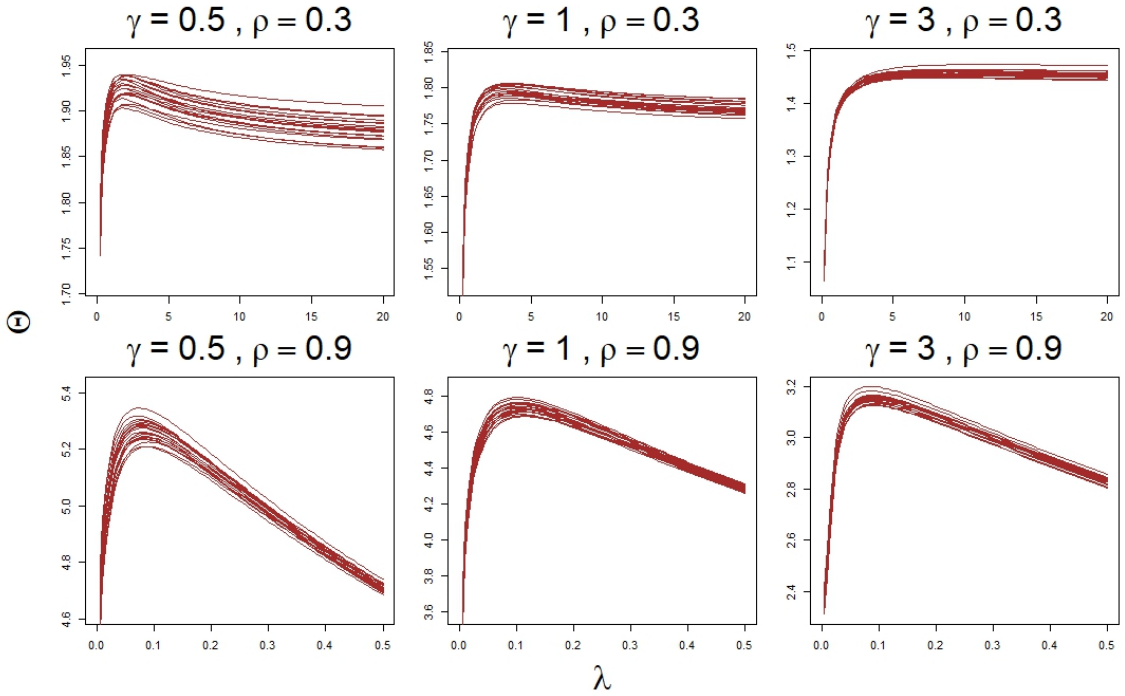

Figure 2

4.3 AR(1)

4.3.1 Plot of $\Theta(\lambda)$

The plot in Figure 2 show the relationship between Θ and λ for distinct values of $\gamma = 0.5, 1, 3$ under **AR(1)** covariance. For each plot, it is a result of 20 generation of Y and X respectively. The shape of these graphs exhibit an increase followed by decrease in general. However, the degree of the decrease is different depending on the value of γ. As γ becomes bigger, the degree of the decrease after the maximum of Θ gets smaller. When the value ρ is even small, the shape of the plot is similar to nondecreasing function.

Around the left region of given λ, the Θ attains maximum for each given γ and ρ, where the optimal lambda λ_{opt} occurs. Throughout this figure, it is seen that the λ_{opt} tends to increase as γ increases for fixed ρ. Also, λ_{opt} decreases as the value of ρ increases for the same γ.

4.3.2 λ_{opt}

This section covers the relationship between λ_{opt} vs γ for different $\rho = 0.3, 0.5, 0.7, 0.9$. Each λ_{opt} is obtained by finding the value of λ at which the $\Theta(\lambda)$ attains maximum for given γ.

The Figure 3 shows how the λ_{opt} changes as γ increases. The λ_{opt} tends to increase as the γ increases in common regardless the value of ρ. For a fixed γ, the value of λ_{opt} tends to decrease as the value of ρ increases. For the case of high correlation $\rho = 0.9$, there is a region of γ where the λ_{opt} decreases, similar to a cubic function.

4.3.3 $\Theta(\lambda_{opt}), \Theta(\infty)$ for $\gamma > 0$

In this section, we'll compare the outline of $\Theta(\lambda_{opt})$ with the case $\lambda = \infty$, which corresponds to the mean difference. $\Theta(\infty)$ is a simplified notation of

$$\Theta(\infty) = \lim_{\lambda \to \infty} \Theta(\lambda).$$

This is valid for all $\gamma > 0$ according to Dobriban and Wager (2018). Recall that the convergence $\lim_{\lambda\to\infty} m(-\lambda), \lim_{\lambda\to\infty} v(-\lambda)$ is valid for all $\gamma > 0$, according to the proof of **Theorem 3.5** of

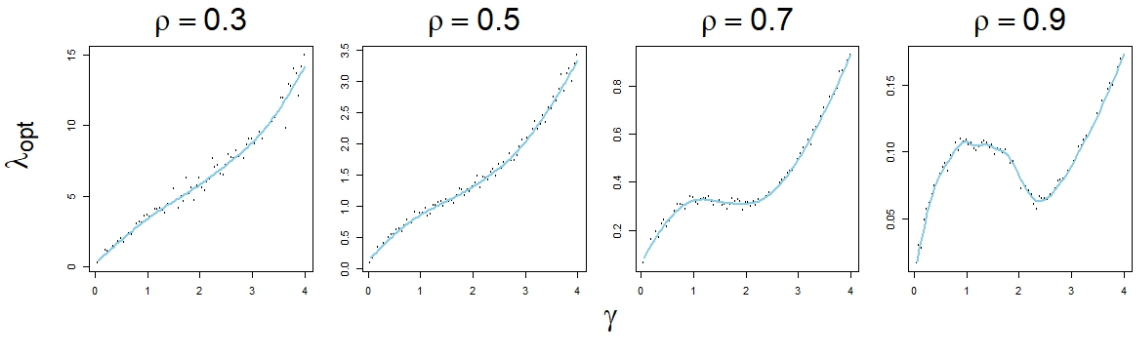

Figure 3: Optimal λ vs γ

Dobriban and Wager (2018). However, the limit $\lim_{\lambda\to 0} m(-\lambda)$ and $\lim_{\lambda\to 0} v(-\lambda)$ under $\gamma > 1$ is not guaranteed, based on the same theorem. So, we compare $\Theta(\lambda_{opt})$ and $\Theta(\infty)$ only under all $\gamma > 0$.

Despite the above definitions, we cannot directly obtain the value $\Theta(\infty)$ when using the almost sure approximation. Going back to the definition of $\Theta(\lambda)$ in (22), it consists of τ, η, and ξ which are functions of m, v. Using the almost sure convergence, each $m(z), v(z)$ and their derivatives can be approximated as (using (28))

$$m(z) \approx \frac{1}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j - z}, \quad v(z) \approx \gamma\left(\frac{1}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j - z} + \frac{1}{z}\right) - \frac{1}{z}. \tag{31}$$

$$m'(z) \approx \frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2}, \quad v'(z) \approx \gamma\left(\frac{1}{p}\sum_{j=1}^{p}\frac{1}{(\lambda_j - z)^2} - \frac{1}{z^2}\right) + \frac{1}{z^2} \tag{32}$$

Switching $-z := \lambda$, we have

$$m(-\lambda) \asymp 1/\lambda, \quad v(-\lambda) \asymp 1/\lambda \tag{33}$$

$$m'(-\lambda) \asymp 1/\lambda^2, \quad v'(-\lambda) \asymp 1/\lambda^2. \tag{34}$$

for some sufficiently large $\lambda > 0$. Then, the expression of $\Theta(\lambda)$ in (22) becomes

$$\Theta(\lambda) = \frac{\alpha^2\tau(\lambda)}{\sqrt{\alpha^2\eta(\lambda) + \xi(\lambda)}} \tag{35}$$

$$= \alpha^2 \cdot \frac{\lambda m v^2}{\sqrt{\alpha^2(v^3 - \lambda v' v^2)/\gamma + (v' - v^2)}} \tag{36}$$

$$\lesssim \frac{\lambda \cdot (1/\lambda^3)}{\sqrt{\alpha^2(v^3 - \lambda v' v^2)/\gamma + (v' - v^2)}} \tag{37}$$

$$\lesssim \frac{\lambda \cdot (1/\lambda^2)}{\sqrt{1/\lambda^3 + 1/\lambda^2}} \tag{38}$$

$$\leq \frac{(1/\lambda^2)}{\sqrt{1/\lambda^2}} \tag{39}$$

$$= \frac{1}{\lambda} \tag{40}$$

For sufficiently large $\lambda > 0$. Since this results the approxmation of $\Theta(\infty)$ in 0, the λ should not be infinitely large when empirically calculating $\Theta(\infty)$. Empirically, we obtained $\lambda \in [10^5, 10^6]$ produces stable approximation of $\Theta(\infty)$.

The Figure 4 shows how the logarithm of $\Phi(-\Theta)$ changes as $\gamma > 0$ increases for $\lambda = \lambda_{opt}$ and ∞. The red graph is the logarithm of $\Phi(-\Theta(\lambda_{opt}))$ for each of γ, and the green graph is $\Phi(-\Theta(\infty))$ at given γ respectively. Regardless of the value of ρ, all of the $\log\Phi(-\Theta)$ tend to increase as γ increases, and the two graphs are getting closer as $\gamma \to \infty$. The difference is the degree of the disparity between λ_{opt} and $\lambda = \infty$. As the ρ increases, the difference in logarithm of $\Phi(-\Theta)$ increases as well.

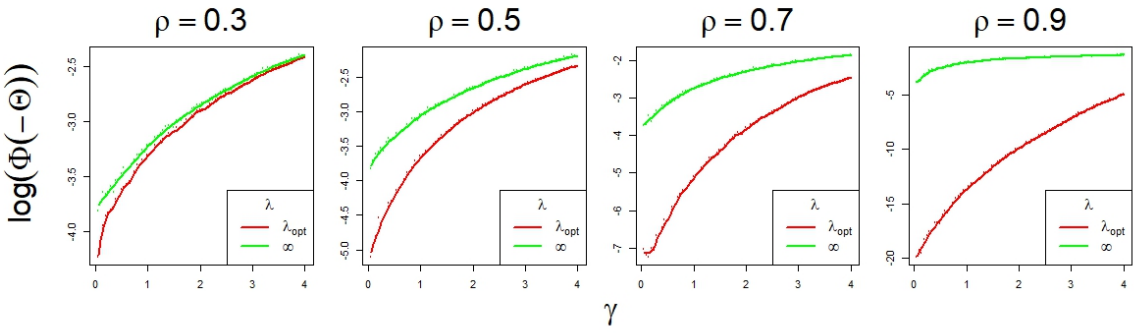

Figure 4

4.3.4 $\Theta(\lambda_{opt}), \Theta(\infty), \Theta(0)$ for $0 < \gamma < 1$

In this section, we'll compare all three values of $\Theta(\lambda_{opt}), \Theta(\infty), \Theta(0)$. The $\Theta(0)$ is a simiplification of

$$\Theta(0) = \lim_{\lambda \to 0} \Theta(\lambda). \tag{41}$$

The article Dobriban and Wager (2018) showed the convergence of $m(-\lambda), v(-\lambda)$ is guaranteed only in the region $0 < \gamma < 1$. So, we compare Θ on the region $0 < \gamma < 1$.

In order to check the value of $\Theta(0)$, we cover the case $\gamma > 1$ first. The eigenvalues λ_j is positive for $j = 1, \ldots, n$ while $\lambda_j = 0$ for $n + 1 \le j \le p$. Then, we have same results as in the (33) and (34) as $\lambda \to 0$. Then,

$$\begin{aligned}
\Theta(\lambda) &= \alpha^2 \cdot \frac{\lambda m v^2}{\sqrt{\alpha^2(v^3 - \lambda v' v^2)/\gamma + (v' - v^2)}} && (42)\\
&\gtrsim \frac{\lambda \cdot (1/\lambda^3)}{\sqrt{\alpha^2(v^3 - \lambda v' v^2)/\gamma + (v' - v^2)}} && (43)\\
&\gtrsim \frac{\lambda \cdot (1/\lambda^3)}{\sqrt{1/\lambda^3 + 1/\lambda^2}} && (44)\\
&\ge \frac{\lambda \cdot (1/\lambda^3)}{\sqrt{1/\lambda^3}} && (45)\\
&\ge \frac{1}{\lambda^{1/2}} && (46)
\end{aligned}$$

for sufficiently small $\lambda > 0$, which makes $\Theta(0) = \infty$.

When $\gamma < 1$, this implies all the eigenvalues to be positive (assuming full column rank of X). Then, the approximation of $m(-\lambda)$ in (26) becomes the average of the reciprocal of all eigenvalues (contant). However, the approximation of $v(-\lambda)$ in (27) still has the form $1/\lambda$, so

$$v(-\lambda) \asymp 1/\lambda, \qquad v'(-\lambda) \asymp 1/\lambda^2 \tag{47}$$

still holds in $0 < \gamma < 1$ as well. Then,

$$\Theta(\lambda) \asymp \frac{\lambda \cdot (1/\lambda^2)}{\sqrt{1/\lambda^3 + 1/\lambda^2}} \asymp \frac{1/\lambda}{\sqrt{1/\lambda^3}} = \sqrt{\lambda} \tag{48}$$

which makes $\Theta(0) = 0$.

To sum up, the if the almost sure convergence is used for the approximation of m, v, then $\Theta(0)$ becomes ∞ for $\gamma > 1$ and 0 for $\gamma < 1$, regardless of real value of $\Theta(0)$. So, we need some sufficiently small value λ that returns stable value of $\Theta(0)$. Empirically, $\lambda \in [10^{-3}, 10^{-2}]$ produces stable value for $\Theta(0)$.

The Figure 5 shows how the logarithm of $\Phi(-\Theta)$ changes in the range of $\gamma \in (0,1)$ for $\lambda = \lambda_{opt}, 0.$ and ∞. It can be shown that the $\log \Phi(-\Theta)$ of $\lambda = \lambda_{opt}, 0$ is almost close when $\gamma = 0$. As ρ increases, the $\Phi(-\Theta(\infty))$ becomes the largest among others. Also, the γ where $\Phi(-\Theta(0))$ and $\Phi(-\Theta(\infty))$ become equal get to increase as ρ increases.

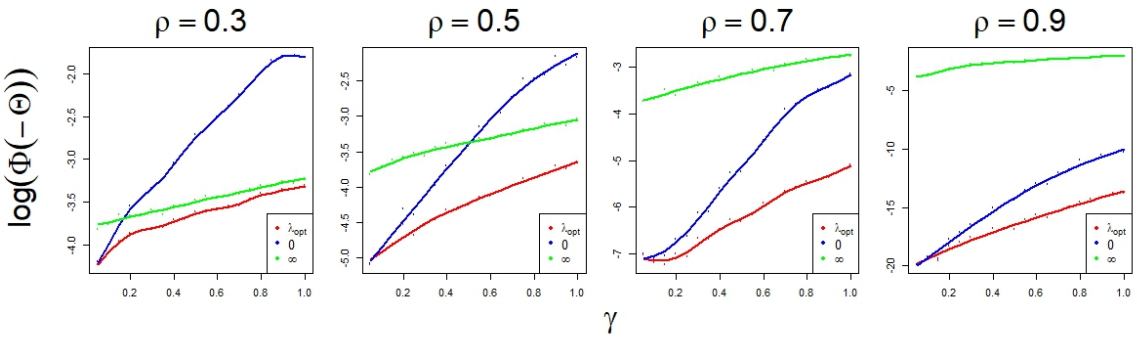

Figure 5

In summary, in AR(1) covariance structure, $\lambda = \infty$ (Mean Difference) generates lower classification error at higher dimension, while $\lambda = 0$ (MDP) works better as the dimension gets smaller. The phenomenon of the $\Phi(-\Theta)$ having biggest value around $\gamma = 1$ is similar to "fig.2" of Ahn and Marron (2010), which also shows that the classification errors of MDP is getting smaller than that of MD under compound symmetry. As previously mentioned, this interpretation is valid only for $\gamma < 1$.

Summary of empirical patterns (AR(1)). Across $\rho \in \{0.3, 0.5, 0.7, 0.9\}$ and $\gamma \in (0, 4]$:

- λ_{opt} increases monotonically with γ (except for mild nonmonotonicity at very high correlation, $\rho = 0.9$, where a shallow local dip appears).
- For fixed γ, λ_{opt} decreases as ρ increases.
- $\Phi(-\Theta(\lambda_{\text{opt}}))$ approaches $\Phi(-\Theta(\infty))$ as γ grows, whereas for small γ it is close to $\Phi(-\Theta(0))$.

Table 1 summarizes these monotonicities.

Table 1: AR(1): qualitative relations observed across all settings.

Quantity	$\gamma \uparrow$	$\rho \uparrow$
λ_{opt}	increases	decreases
$\lvert\Theta(\lambda_{\text{opt}}) - \Theta(\infty)\rvert$	decreases	increases
$\lvert\Theta(\lambda_{\text{opt}}) - \Theta(0)\rvert$	increases	decreases

4.4 Matérn Covariance

Until previous sections, we considered the case where Σ is AR(1). From now on, the covariance is set to "Matérn Covariance" function :

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{d}{\rho}\right),$$

where ν, ρ are positive parameters, d is a distance between ith and j coordinates (i.e. $d = |i-j|$), and K_ν is a modified Bessel function. When ν has specific value, such as $\nu = n + 1/2, n \in \mathbb{Z}^+$, the Matérn Covariance function also has simpler form. For example,

$$C_{1/2} = \sigma^2 \exp\left\{-\frac{d}{\rho}\right\},$$

$$C_{3/2} = \sigma^2\left(1+\frac{\sqrt{3}d}{\rho}\right)\exp\left\{-\frac{\sqrt{3}d}{\rho}\right\},$$
$$C_{5/2} = \sigma^2\left(1+\frac{\sqrt{5}d}{\rho}+\frac{5d^2}{3\rho^2}\right)\exp\left\{-\frac{\sqrt{5}d}{\rho}\right\}$$

and

$$\lim_{\nu\to\infty} C_\nu(d) = \sigma^2\exp\left\{-\frac{d^2}{2\rho^2}\right\}.$$

In this section, we'll use the simplest case $\nu = 1/2$.

4.4.1 Plot of $\Theta(\lambda)$

The Figure 6 shows the plot of $\Theta(\lambda)$ for distinct ρ's. As in the case of AR(1), each plot is a result of 20 generations of $\Theta(\lambda)$. In the case of $\rho = 0.9$, the trend of increase followed by decrease is stark. Although the similar trend exists in the case $\rho = 0.5$, the degree of decrease after attaining maximum of Θ is extremely low.

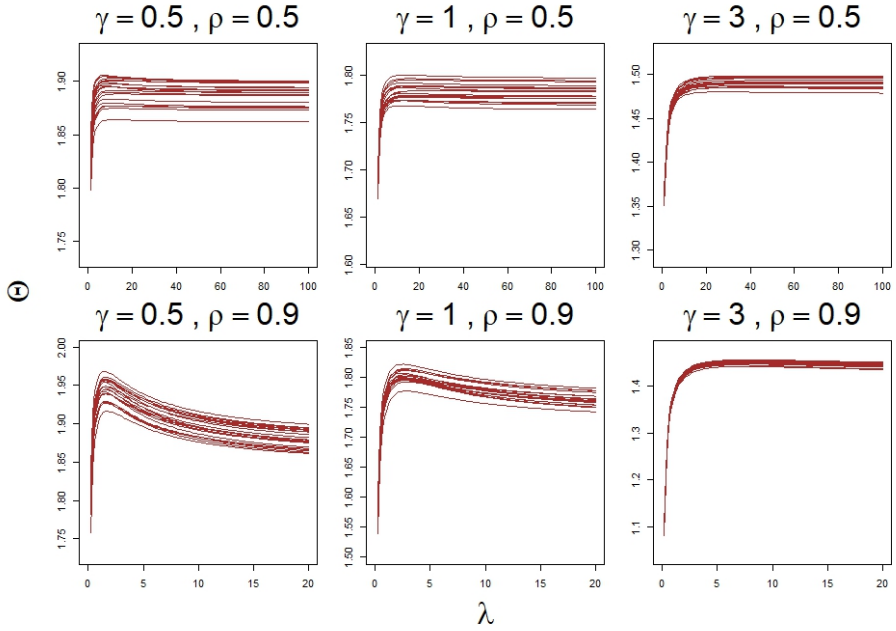

Figure 6

4.4.2 λ_{opt}

Following Figure 7 shows the relationship between λ_{opt} vs γ for different values of $\rho = 0.5, 0.7, 0.9, 0.99$. Commonly, the optimal λ grows gradually as the γ increases, similar to the case of AR(1). For a fixed γ, the optimal λ values becomes lower as the ρ increases.

4.4.3 $\Theta(\lambda_{opt}), \Theta(\infty)$ for $\gamma > 0$

The Figure 8 shows graphs of the logarithm of $\Phi(-\Theta)$ vs γ in different ρ values, where $\lambda = \lambda_{opt}, \infty$. When ρ is relatively low (0.5), the Θ at $\lambda = \lambda_{opt}$ and $\lambda = \infty$ are almost similar. This complies with the case $\gamma = 3, \rho = 0.5$ of the Figure 6 where the Θ value is almost similar after sufficiently large $\lambda > 0$. The general trend of the graphs, such as the increase of $\Phi(-\Theta)$ as γ increases or the fact that $\Phi(-\Theta)$ is greater in the case of $\lambda = \infty$ than in the case $\lambda = \lambda_{opt}$, is similar to that of AR(1). However, the differene of $\Phi(-\Theta)$ between the case $\lambda = \infty$ and λ_{opt} is not significant, even after the change into logarithm.

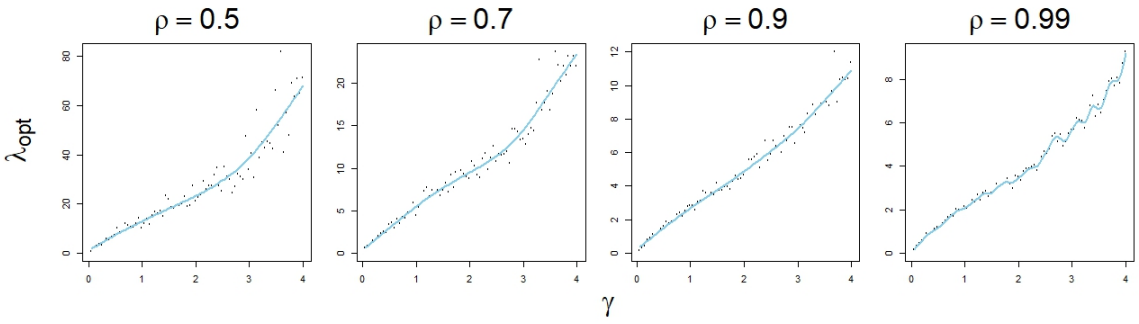

Figure 7: Optimal λ, Matern Covariance

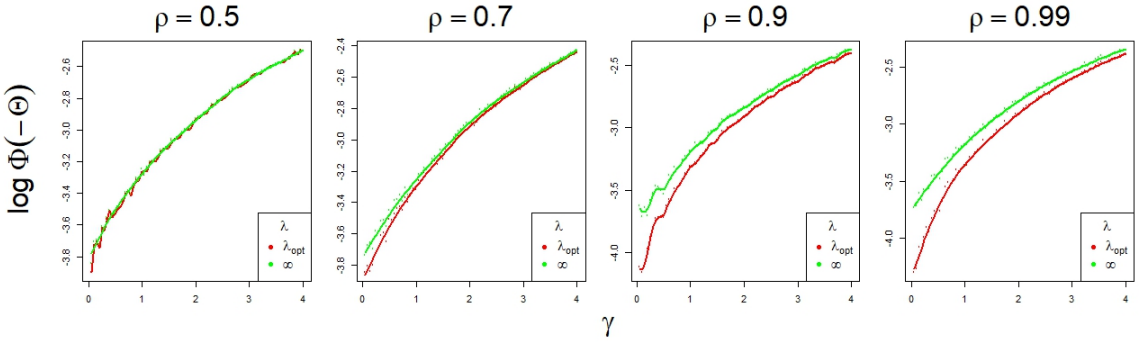

Figure 8

4.4.4 $\Theta(\lambda_{opt}), \Theta(\infty), \Theta(0)$ for $0 < \gamma < 1$

The Figure 9 shows $\log \Phi(-\Theta)$ vs $\gamma \in (0,1)$ where $\lambda = \lambda_{opt}, 0, \infty$ for different ρ. In common, the optimal Θ is close to $\Theta(0)$ when $\gamma = 0$, while $\Theta(\lambda_{opt})$ is getting close to $\Theta(\infty)$ as γ increases. In summary, $\lambda = 0$ (MDP) performs better in low dimension, while $\lambda = \infty$ (MD) works better as the dimension gets larger in Matern Covariance as well.

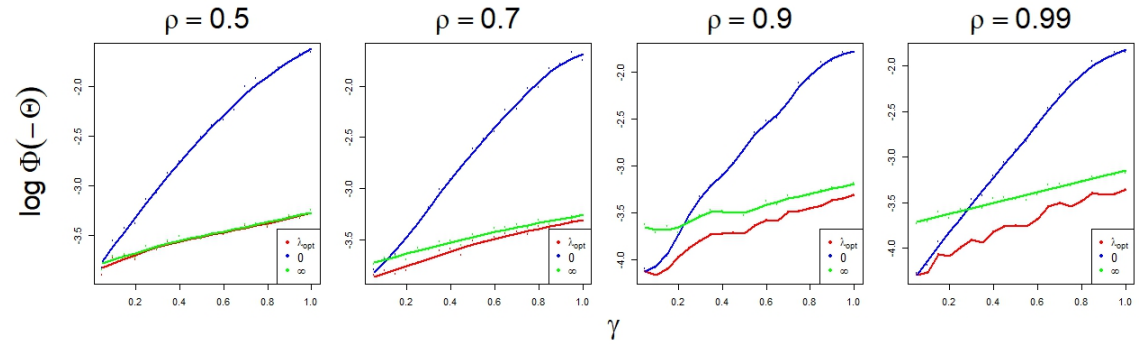

Figure 9

Summary of empirical patterns (Matérn, $\nu = \frac{1}{2}$). The trends mirror AR(1): λ_{opt} grows with γ and shrinks with ρ. A notable difference is that the gap between $\Phi(-\Theta(\lambda_{\text{opt}}))$ and $\Phi(-\Theta(\infty))$ is uniformly smaller under Matérn for the same (γ, ρ), reflecting stronger effective smoothing of high–frequency directions. This explains why the red/green curves in Fig. 8 nearly coincide when ρ is moderate (e.g., $\rho = 0.5$).

4.5 Comparison with prior work

Lee, Ahn & Jeon (2013). They established that $v(\lambda) = (CC^\top + \lambda I)^{-1}w$ forms a continuous path connecting $v(0) = \text{MDP}$ and $v(\infty) = \text{MD}$. Building on the limiting risk formula of Dobriban and Wager (2018), we track how the error behaves along this path through $\Theta(\lambda)$ and, in particular, document the transition rules (section 4):

1. λ_{opt} increases as γ grows,
2. λ_{opt} decreases as correlation strengthens (AR range or Matérn smoothness/range),
3. $\Theta(\lambda_{\text{opt}})$ is close to $\Theta(0)$ at small γ and close to $\Theta(\infty)$ at large γ.

5 Conclusion

This paper examined regularized discriminant analysis in the HDLSS regime through the random matrix theory, connecting the MDP–MD continuum to asymptotic misclassification risk. Our analysis makes the limiting error formula $\text{Err}(\hat{w}_\lambda) \to \Phi\big(-\Theta(\lambda)\big)$ explicit and computable in settings of practical interest, and clarifies when the extremes $\lambda = 0$ (MDP) and $\lambda = \infty$ (MD) are preferable.

Summary of main findings. (i) In Section 3, we derived a piecewise closed form for the MDP limit $\Theta(0)$ under $\Sigma = I$ across the regimes $\gamma < 1$, $\gamma = 1$, and $\gamma > 1$, which produces a theoretical error curve $\Phi(-\Theta(0))$ with a peak at $\gamma = 1$ and reconciles classical HDLSS observations. (ii) In Section 2 and 4, we verified via Szegő arguments that Toeplitz families (including AR(1) and Matérn) satisfy the spectral assumptions required by the limiting risk, thereby justifying their use in our computations. (iii) In Section 4, we developed a numerically stable approximation pipeline for $\Theta(\lambda)$ and mapped the optimizer $\lambda_{\text{opt}}(\gamma, \rho)$. Empirically, λ_{opt} *increases* with γ and *decreases* with correlation strength ρ; moreover, $\Theta(\lambda_{\text{opt}})$ is close to $\Theta(0)$ at small γ and close to $\Theta(\infty)$ at large γ, quantifying the transition between MDP and MD along the regularization path.

Theoretical contribution. The results turn the abstract Stieltjes-transform limit for rLDA into closed-form and computable objects in representative covariance models. The $\Sigma = I$ formula provides a clean baseline for the MDP end, while the Toeplitz verification extends applicability to structured dependence and grounds the numerical evaluation of $\Theta(\lambda)$ and λ_{opt}.

Practical implications. The maps $(\gamma, \rho) \mapsto \lambda_{\text{opt}}$ offer simple guidance for tuning: when the aspect ratio γ is small, near-unregularized directions (MDP-like) are competitive; as γ grows or correlation weakens, stronger ridge regularization is beneficial. In highly correlated designs (large ρ), the gap between $\Phi\big(-\Theta(\lambda_{\text{opt}})\big)$ and $\Phi\big(-\Theta(\infty)\big)$ narrows, indicating that a mean-difference rule can be near-optimal. These patterns help practitioners choose λ before resorting to cross-validation.

Limitations. Our study focuses on balanced binary classification, Gaussian data with covariance structures in the Toeplitz class, and ridge-type regularization. The $\gamma = 1$ case is singular for the a.s. approximations and is excluded from grids. Endpoint evaluations rely on stable proxies $\lambda_0 \in [10^{-3}, 10^{-2}]$ and $\lambda_\infty \in [10^5, 10^6]$, which are numerically motivated.

Future directions. Several natural extensions remain.

- **Mild class imbalance and heteroscedasticity.** Extend the derivations to $\Pr(Y = 1) = \pi \neq \frac{1}{2}$ and class-specific Toeplitz covariances $\Sigma_+ \neq \Sigma_-$, re-expressing $\Theta(\lambda)$ with an adjusted margin scale and pooled vs. discriminant covariance choices; empirically check whether the monotonicities of λ_{opt} in (γ, ρ) persist.
- **Near the singular point $\gamma \approx 1$.** Replace the hard exclusion by a controlled neighborhood with a small analytic ridge (e.g., $\widehat{\Sigma} + \varepsilon I$) and a matched-asymptotic expansion of $\Theta(\lambda)$; compare to our proxy curves to quantify the approximation error.
- **Beyond Gaussian but still light-tailed.** Replace Gaussian Z with sub-Gaussian or elliptical models with finite fourth moment and re-verify the a.s. convergence in our pipeline.

6 Appendix

The $\Theta(0)$ varies depending on the value of $\gamma > 0$.

6.1 $\gamma < 1$

We'll use such notation as $\tau(0) := \lim_{\lambda \to 0} \tau(\lambda)$ to simplify it.

(1) $\boldsymbol{\tau(0)}$

Since

$$m(z) = \frac{4\gamma z}{2\gamma z} \frac{1}{1-\gamma-z+\sqrt{(1-\gamma-z)^2}}, \tag{49}$$

we have

$$m(0) = \frac{1}{1-\gamma}. \tag{50}$$

Using (19), we conclude

$$zv(z) = \gamma z m(z) + (\gamma - 1) \tag{51}$$
$$zv(z)m(z) = \gamma z m^2(z) + (\gamma - 1)m(z) \tag{52}$$
$$\lambda v(-\lambda)m(-\lambda) = \gamma \lambda m^2(-\lambda) + (1-\gamma)m(-\lambda) \tag{53}$$
$$\tau(0) = (1-\gamma)/(1-\gamma) = 1. \tag{54}$$

(2) $\boldsymbol{\eta(0)}$

Using the definition of v in (19),

$$v'(z) = \gamma m'(z) - \frac{\gamma - 1}{z^2}. \tag{55}$$

Then,

$$\eta(z) = \frac{1}{\gamma}\left(\gamma m(z) + \frac{\gamma-1}{z} + z\gamma m'(z) - \frac{\gamma-1}{z}\right) = \frac{1}{\gamma}\Big(\gamma m(z) + z\gamma m'(z)\Big). \tag{56}$$

Now, the remaining task is to calculate $m'(0)$.

$$m'(z) = \frac{\big(1-\gamma-z-\sqrt{(1-\gamma-z)^2-4\gamma z}\big)'(2\gamma z) - \big(1-\gamma-z-\sqrt{(1-\gamma-z)^2-4\gamma z}\big)(2\gamma)}{(2\gamma z)^2} \tag{57}$$
$$= \frac{\left(-1-\frac{1}{2}\frac{2(z+\gamma-1)-4\gamma}{\sqrt{(1-\gamma-z)^2-4\gamma z}}\right)(2\gamma z) - \big(1-\gamma-z-\sqrt{(1-\gamma-z)^2-4\gamma z}\big)(2\gamma)}{(2\gamma z)^2} \tag{58}$$
$$= \frac{-\frac{2(z+\gamma-1)-4\gamma}{\sqrt{(1-\gamma-z)^2-4\gamma z}}(\gamma z) - 2\gamma(1-\gamma) + 2\gamma\sqrt{(1-\gamma-z)^2-4\gamma z}}{(2\gamma z)^2}. \tag{59}$$

Using L'Hopital's rule,

$$m'(0) = \frac{-z\left\{\frac{2}{\sqrt{(1-\gamma-z)^2-4\gamma z}} - \frac{\big(2(z+\gamma-1)-4\gamma\big)^2}{2\big(\sqrt{(1-\gamma-z)^2-4\gamma z}\big)^3}\right\}}{8\gamma^2 z} \tag{60}$$
$$= \frac{-1}{8\gamma^2}\left\{\frac{2}{1-\gamma} - \frac{(-2\gamma-2)^2}{2(1-\gamma)^3}\right\} \tag{61}$$
$$= \frac{1}{\gamma(1-\gamma)^2}. \tag{62}$$

Going back to the equation 56,

$$\eta(0) = m(0) + 0 \cdot m'(0) = \frac{1}{1-\gamma}. \tag{63}$$

(3) $\boldsymbol{\xi(0)}$

$$\frac{v'(z)}{v^2(z)} = \frac{\gamma m'(z) - \dfrac{\gamma-1}{z^2}}{\gamma^2 m^2(z) + 2\dfrac{m(z)}{z}\gamma(\gamma-1) + \dfrac{(\gamma-1)^2}{z^2}} \tag{64}$$

$$= \frac{z^2\gamma m'(z) - (\gamma-1)}{z^2\gamma^2 m^2(z) + 2zm(z)\gamma(\gamma-1) + (\gamma-1)^2} \tag{65}$$

Using the results (50) and (62),

$$\frac{v'(0)}{v^2(0)} = \frac{-1}{\gamma-1} \tag{66}$$

and

$$\xi(0) = \frac{\gamma}{\gamma-1}. \tag{67}$$

(4) conlcusion

Combining (25), (54), (63), (67) together, this results in

$$\Theta(0) = \frac{\alpha^2}{\sqrt{\alpha^2\dfrac{1}{1-\gamma} + \dfrac{\gamma}{1-\gamma}}}. \tag{68}$$

6.2 $\gamma = 1$

Applying $\gamma = 1$ in (20) and (21), we have $m(z) = v(z)$ for $\gamma = 1$. So,

$$m(z) = -\frac{1}{2} - \frac{1}{2}\sqrt{1-\frac{4}{z}} \asymp -\frac{1}{\sqrt{-z}} \tag{69}$$

$$\Rightarrow m(-\lambda) \asymp -\frac{1}{\sqrt{\lambda}}. \tag{70}$$

The equation (59) can be used to obtain $m'(z)$ at $\gamma = 1$.

$$m'(z) = \frac{-\dfrac{2(z-2)}{\sqrt{z^2-4z}}z + 2\sqrt{z^2-4z}}{4z^2} \tag{71}$$

$$= -\frac{1}{2\sqrt{z^2-4z}} + \frac{1}{z\sqrt{z^2-4z}} + \frac{\sqrt{z^2-4z}}{2z} \tag{72}$$

$$\asymp (-z)^{-3/2} \tag{73}$$

$$\Rightarrow m'(-\lambda) \asymp \lambda^{-3/2}. \tag{74}$$

Hence, we can obtain

$$\tau(\lambda) \asymp 1 \tag{75}$$

$$\eta(\lambda) \asymp -\frac{1}{\sqrt{\lambda}} - \lambda(\lambda)^{-3/2} \asymp -\frac{1}{\sqrt{\lambda}} \tag{76}$$

$$\xi(\lambda) \asymp \frac{\lambda^{-3/2}}{1/\lambda} \asymp \lambda^{-1/2} \tag{77}$$

$$\alpha^2\eta(\lambda) + \xi(\lambda) \asymp \lambda^{-1/2} \tag{78}$$

$$\Theta(\lambda) \asymp \lambda^{1/4} \tag{79}$$

$$\Rightarrow \Theta(0) = 0. \tag{80}$$

6.3 $\gamma > 1$

(1) $\boldsymbol{\tau(0)}$

We may use the definition of companion Stieltjes transform (19) to obtain

$$m(z) = \frac{1}{\gamma}\bigg(v(z) + \frac{1-\gamma}{z}\bigg). \tag{81}$$

Then,

$$zm(z)v(z) = \frac{1}{\gamma}\bigg(zv(z) + (1-\gamma)\bigg)v(z). \tag{82}$$

The $v(z)$ can be expressed as

$$\begin{aligned} v(z) &= \frac{\gamma - 1 - z - \sqrt{(\gamma-1+z)^2 - 4\gamma z}}{2z} && (83)\\ &= \frac{(\gamma-1-z)^2 - (\gamma-1+z)^2 + 4\gamma z}{2z\big(\gamma - 1 - z + \sqrt{(\gamma-1+z)^2 - 4\gamma z}\big)} && (84)\\ &= \frac{-4(\gamma-1)z + 4\gamma z}{2z\big(\gamma - 1 - z + \sqrt{(\gamma-1+z)^2 - 4\gamma z}\big)} && (85)\\ &= \frac{2}{\gamma - 1 - z + \sqrt{(\gamma-1+z)^2 - 4\gamma z}}. && (86) \end{aligned}$$

So, $v(0) = \frac{2}{2(\gamma-1)}$ and thus

$$\tau(0) = -\frac{1-\gamma}{\gamma}\frac{1}{\gamma-1} = \frac{1}{\gamma}. \tag{87}$$

(2) $\boldsymbol{\eta(0)}$

$$\begin{aligned} v'(z) &= \frac{\bigg(-1 - \dfrac{2(z+\gamma-1)-4\gamma}{2\sqrt{(1-\gamma-z)^2 - 4\gamma z}}\bigg)2z - 2\big(\gamma - 1 - z - \sqrt{(1-\gamma-z)^2 - 4\gamma z}\big)}{4z^2} && (88)\\ &= \frac{-z\dfrac{z-\gamma-1}{\sqrt{(1-\gamma-z)^2 - 4\gamma z}} - (\gamma-1) + \sqrt{(1-\gamma-z)^2 - 4\gamma z}}{2z^2}. && (89) \end{aligned}$$

Applying L'Hopital's rule results in

$$\begin{aligned} v'(0) &= -\frac{1}{4}\bigg\{\frac{1}{(1-\gamma)^2}\bigg(\sqrt{(1-\gamma)^2} + (\gamma+1)\frac{-\gamma-1}{\sqrt{(1-\gamma)^2}}\bigg)\bigg\} && (90)\\ &= -\frac{1}{4}\frac{1}{(1-\gamma)^2}\bigg((\gamma-1) - \frac{(\gamma+1)^2}{\gamma-1}\bigg) && (91)\\ &= \frac{\gamma}{(\gamma-1)^3}. && (92) \end{aligned}$$

Thus

$$\eta(0) = \frac{1}{\gamma}\big(v(0) + 0\cdot v'(0)\big) = \frac{1}{\gamma}v(0) = \frac{1}{\gamma(\gamma-1)}. \tag{93}$$

(3) $\boldsymbol{\xi(0)}$

$$\xi(0) = \frac{\dfrac{\gamma}{(\gamma-1)^3}}{\dfrac{1}{(\gamma-1)^2}} - 1 = \frac{1}{\gamma-1}. \tag{94}$$

(4) conclusion

$$\Theta(0) = \frac{\alpha^2/\gamma}{\sqrt{\alpha^2\dfrac{1}{\gamma(\gamma-1)} + \dfrac{1}{\gamma-1}}} = \frac{\alpha^2}{\sqrt{\alpha^2\dfrac{\gamma}{\gamma-1} + \dfrac{\gamma^2}{\gamma-1}}}. \tag{95}$$

This shows how the result of Corollary 1 is derived.

References

JeongYoun Ahn and J.S. Marron. The maximal data piling direction for discrimination. *Biometrika*, 2010.

Z. Bai and J. W. Silverstein. Spectral analysis of large dimensional random matrices. *Springer*, 2010.

E. Dobriban and S. Wager. High-dimensional asymptotics of prediction: Ridge regression and classification. *The Annals of Statistics*, 2018.

J.W.Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 1995.

Myung Hee Lee, Jeongyoun Ahn, and Yongho Jeon. Hdlss discrimination with adaptive data piling. *Journal of Computational and Graphical Statistics*, 2013.

Marchenko and Pastur. Distribution of eigenvalues for some sets of random matrices. *American Mathematical Society*, 1967.