

Notes for High-Dimensional Probability Second Edition by
Roman Vershynin

Gallant Tsao

July 24, 2025

Contents

0	Appetizer: Using Probability to Cover a Set	4
0.1	Covering Geometric Sets	5
1	A Quick Refresher on Analysis and Probability	7
1.1	Convex Sets and Functions	7
1.2	Norms and Inner Products	7
1.3	Random Variables and Random Vectors	7
1.4	Union Bound	8
1.5	Conditioning	9
1.6	Probabilistic Inequalities	9
1.7	Limit Theorems	11
2	Concentration of Sums of Independent Random Variables	13
2.1	Why Concentration Inequalities?	13
2.2	Hoeffding Inequality	14
2.3	Chernoff Inequality	16
2.4	Application: Median-of-means Estimator	18
2.5	Application: Degrees of Random Graphs	20
2.6	Subgaussian Distributions	21
2.6.1	The Subgaussian Norm	23
2.7	Subgaussian Hoeffding and Khintchine Inequalities	23
2.7.1	Subgaussian Hoeffding Inequality	24
2.7.2	Subgaussian Khintchine Inequality	24
2.7.3	Maximum of Subgaussians	25
2.7.4	Centering	26
2.8	Subexponential Distributions	26
2.8.1	Subexponential Properties	26
2.8.2	The Subexponential Norm	28
2.9	Bernstein Inequality	29
3	Random Vectors in High Dimensions	32
3.1	Concentration of the Norm	32
3.2	Covariance Matrices and PCA	33
3.2.1	Learning from the Covariance Matrix	33
3.2.2	Principle Component Analysis	34
3.2.3	Isotropic Distributions	35
3.3	Examples of High-dimensional Distributions	35
3.3.1	Standard Normal	35
3.3.2	General Normal	36
3.3.3	Uniform on the Sphere	37
3.3.4	Uniform on a Convex Set	38
3.3.5	Frames	38
3.4	Subgaussian Distributions in High Dimensions	40
3.4.1	Gaussian, Rademacher, and More	40
3.4.2	Uniform on the Sphere	40
3.4.3	Non-examples	41
3.5	Application: Grothendieck Inequality and Semidefinite Programming	42
3.5.1	Semidefinite Programming	44
3.6	Application: Maximum Cut for Graphs	46
3.6.1	A Simple 0.5-approximation Algorithm	46
3.6.2	Semidefinite Relaxation	47
3.7	Kernel Trick and Tightening of Grothendieck Inequality	48
3.7.1	Tensors	49
3.7.2	Proof of Theorem 3.5.1	51
3.7.3	Kernels and Feature Maps	51

4	Random Matrices	52
4.1	A Quick Refresher on Linear Algebra	52
4.1.1	Singular Value Decomposition	52
4.1.2	Min-max Theorem	53
4.1.3	Frobenius and Operator Norms	54
4.1.4	The Matrix Norms and the Spectrum	54
4.1.5	Low-rank Approximation	55
4.1.6	Perturbation Theory	55
4.1.7	Isometries	57
4.2	Nets, Covering, and Packing	57
4.2.1	Covering Numbers and Volume	59
4.3	Application: Error Correcting Codes	60
4.4	Upper Bounds on Subgaussian Random Matrices	60
4.4.1	Computing the Norm on an ε net	61
4.4.2	The Norms of Subgaussian Random Matrices	61
4.4.3	Symmetric Matrices	63
4.5	Application: Community Detection in Networks	63
4.6	Two-sided Bounds on Subgaussian Matrices	63
4.7	Application: Covariance Estimation and Clustering	65
5	Concentration Without Independence	66
5.1	Concentration of Lipschitz Functions on the Sphere	66
5.1.1	Lipschitz Functions	66
5.1.2	Concentration via Isoperimetric Inequalities	66
5.1.3	Blow-up of Sets on the Sphere	67
5.1.4	Proof of Theorem 5.1.3	68
5.2	Concentration on Other Metric Measure Spaces	69
5.2.1	Gaussian Concentration	69
5.2.2	Hamming Cube	69
5.2.3	Symmetric Group	70
5.2.4	Riemannian Manifolds with Strictly Positive Curvature	70
5.2.5	Special Orthogonal Group	70
5.2.6	Grassmannian	71
5.2.7	Continuous Cube and Euclidean Ball	71
5.2.8	Densities of the Form $e^{-U(x)}$	71
5.2.9	Random Vectors with Independent Bounded Coordinates	72
5.3	Application: Johnson-Lindenstrauss Lemma	72
5.4	Matrix Bernstein Inequality	72
5.4.1	Matrix Calculus	72
5.4.2	Trace Inequalities	74
5.4.3	Proof of Matrix Bernstein Inequality	74
5.4.4	Matrix Hoeffding and Khintchine Inequalities	76
5.5	Application: Community Detection in Sparse Networks	77
5.6	Application: Covariance Estimation for General Distributions	77
5.7	Extra notes	77
6	Quadratic Forms, Symmetrization, and Contraction	78
6.1	Decoupling	78
6.2	Hanson-Wright Inequality	80
6.3	Symmetrization	80
6.4	Random Matrices with non-i.i.d. Entries	81
6.5	Application: Matrix Completion	82
6.6	Contraction Principle	82

7	Random Processes	85
7.1	Basic Concepts and Examples	85
7.1.1	Covariance and Increments	86
7.1.2	Gaussian Processes	86
7.2	Slepian, Sudakov-Fernique, and Gordon Inequalities	87
7.2.1	Gaussian Interpolation	88
7.2.2	Proof of Slepian Inequality	90
7.2.3	Sudakov-Fernique and Gordon Inequalities	91
7.3	Application: Sharp Bounds for Gaussian Matrices	92
7.4	Sudakov Inequality	92
7.4.1	Application for covering numbers in \mathbb{R}^n	93
7.5	Gaussian Width	94
7.5.1	Geometric Meaning of Width	95
7.5.2	Examples	96
7.5.3	Gaussian Complexity and Effective Dimension	97
7.6	Application: Random Projection of Sets	98
8	Chaining	99
8.1	Dudley Inequality	99
8.1.1	Variations and Examples	102
8.2	Application: Empirical Processes	103
8.3	VC Dimension	103
8.3.1	Definition and Examples	103
8.3.2	Pajor's Lemma	105
8.3.3	Sauer-Shelah Lemma	106
8.3.4	Growth Function	106
8.3.5	Covering Numbers via VC Dimension	108
8.3.6	VC Law of Large Numbers	109
8.4	Application: Statistical Learning Theory	112
8.5	Generic Chaining	112
8.5.1	A Makeover of Dudley's Inequality	112
8.5.2	The γ_2 Functional and Generic Chaining	112
8.5.3	Majorizing Measure and Comparison Theorems	114
8.6	Chevet Inequality	116
9	Deviations of Random Matrices on Sets	118
9.1	Matrix Deviation Inequality	118
9.2	Random Matrices, Covariance Estimation, and Johnson-Lindenstrauss	119
9.3	Random Sections: The M^* Bound and Escape Theorem	119
9.4	Application: High-dimensional Linear Models	119
9.5	Application: Exact Sparse Recovery	119
9.6	Deviations of Random Matrices for General Norms	119
9.7	Two-sided Chevet Inequality and Dvoretzky-Milman Theorem	119

9 Deviations of Random Matrices on Sets

The main question in this chapter is: How does an $m \times n$ matrix act on a general set $t \subset \mathbb{R}^n$?

9.1 Matrix Deviation Inequality

Take an $m \times n$ random matrix X with independent, isotropic, and subgaussian rows. The concentration of the norm (Theorem 3.1.1) tells us that for any fixed vector $x \in \mathbb{R}^n$, the approximation

$$\|Ax\|_2 \approx \sqrt{m}\|x\|_2$$

holds with high probability.

Let's ask something more general: Is it true that with high probability, the equation above holds *simultaneously* for many vectors $x \in \mathbb{R}^n$? To quantify how many, pick some bounded set $T \subset \mathbb{R}^n$ and ask if the approximation holds simultaneously for all $x \in T$. It turns out that the maximal error is about $\gamma(T)$, the Gaussian complexity of T .

Theorem 9.1.1 (Matrix deviation inequality). Let A be an $m \times n$ random matrix with independent, isotropic and subgaussian rows A_i . Then for any subset $T \subset \mathbb{R}^n$,

$$\mathbb{E} \left[\sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m}\|x\|_2 \right| \right] \leq CK^2 \gamma(T),$$

where $\gamma(T)$ is the Gaussian complexity from Section 7.5.3, defined as

$$\gamma(T) = \mathbb{E} \left[\sup_{x \in T} |\langle g, x \rangle| \right], \quad g \sim N(0, I_n),$$

and $K = \max_i \|A_i\|_{\psi_2}$.

The plan is to deduce this from Talagrand's comparison inequality (Corollary 8.5.8). To do that, we just have to check the random process

$$Z_x := \|Ax\|_2 - \sqrt{m}\|x\|_2$$

indexed by vectors $x \in \mathbb{R}^n$ has subgaussian increments. Here is the claim:

Theorem 9.1.2 (Subgaussian increments). Let A be an $m \times n$ random matrix with independent, isotropic and subgaussian rows A_i . Then the random process Z_x defined above has subgaussian increments:

$$\|Z_x - Z_y\|_{\psi_2} \leq CK^2 \|x - y\|_2 \text{ for all } x, y \in \mathbb{R}^n,$$

here $K = \max_i \|A_i\|_{\psi_2}$.

Once we have proved this theorem, we plug it into Talagrand's comparison inequality (Exercise 8.37 (a)) and get

$$\mathbb{E} \left[\sup_{x \in T} |Z_x| \right] \leq CK^2 \gamma(T)$$

which directly gives Theorem 9.1.1. So, all we have to do is prove Theorem 9.1.2 - and it is in fact easier since it's for fixed x and y .

Proof of Theorem 9.1.2. This argument will be a bit longer than usual, so we'll (hopefully) make it easier by starting with simpler cases and building up from there.

Step 1: Unit vector x and zero vector y . If $\|x\|_2 = 1$ and $y = 0$, the inequality in the theorem statement becomes

$$\left| \|Ax\|_2 - \sqrt{m} \right|_{\psi_2} \leq CK^2.$$

The random vector $Ax \in \mathbb{R}^m$ has independent, subgaussian coordinates $\langle A_i, x \rangle$, which satisfy

$$\mathbb{E} \left[\langle A_i, x \rangle^2 \right] = 1$$

by isotropy. So, the equation above follows from the concentration of the norm (Theorem 3.1.1).

Step 2: Unit vectors x, y and the squared process. □

- 9.2 Random Matrices, Covariance Estimation, and Johnson-Lindenstrauss
- 9.3 Random Sections: The M^* Bound and Escape Theorem
- 9.4 Application: High-dimensional Linear Models
- 9.5 Application: Exact Sparse Recovery
- 9.6 Deviations of Random Matrices for General Norms
- 9.7 Two-sided Chevet Inequality and Dvoretzky-Milman Theorem