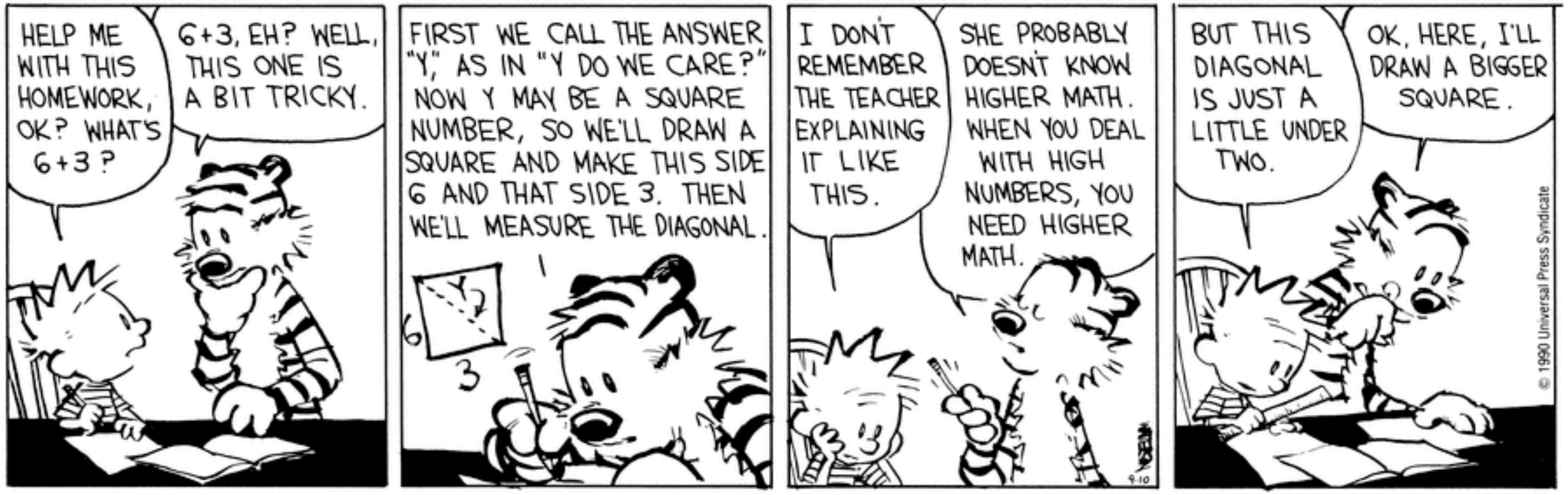
CSC 595 - Research Skills

Math Skills

Nishant Mehta



Mathematical Tricks of the Trade

- Types of things:
  - Cheat sheets
  - Books
  - Tools/Software
  - Techniques (Skills)

Cheat Sheets

[Inequalities cheat sheet]

exponential

$e^x \geq \left(1+\frac{x}{n}\right)^n \geq 1+x; \quad \left(1+\frac{x}{n}\right)^n \geq e^x\left(1-\frac{x^2}{n}\right) \quad \text{for } n \geq 1,\ |x| \leq n.$

$\frac{x^n}{n!}+1 \leq e^x \leq \left(1+\frac{x}{n}\right)^{n+x/2}; \quad e^x \geq \left(\frac{ex}{n}\right)^n \quad \text{for } x, n > 0.$

$x^y + y^x > 1;\ x^y > \frac{x}{x+y};\ e^x > \left(1+\frac{x}{y}\right)^y > e^{\frac{xy}{x+y}};\ \frac{x}{y} \geq e^{\frac{x-y}{x}} \quad \text{for } x, y > 0.$

$\frac{1}{2-x} < x^x < x^2 - x + 1; \quad e^{2x} \leq \frac{1+x}{1-x} \quad \text{for } x \in (0,1).$

$x^{1/r}(x-1) \leq rx(x^{1/r}-1) \quad \text{for } x, r \geq 1; \quad 2^{-x} \leq 1-\frac{x}{2} \quad \text{for } x \in [0,1].$

$xe^x \geq x + x^2 + \frac{x^3}{2}; \quad e^x \leq x + e^{x^2}; \quad e^x + e^{-x} \leq 2e^{x^2/2} \quad \text{for } x \in \mathbb{R}.$

$e^{-x} \leq 1 - \frac{x}{2} \quad \text{for } x \in [0, 1.59]; \quad e^x \leq 1 + x + x^2 \quad \text{for } x < 1.79.$

rearrangement

$$\sum_{i=1}^n a_i b_i \geq \sum_{i=1}^n a_i b_{\pi(i)} \geq \sum_{i=1}^n a_i b_{n-i+1} \quad \text{for } a_1 \leq \cdots \leq a_n,$$

$b_1 \leq \cdots \leq b_n$ and π a permutation of $[n]$. More generally:

$$\sum_{i=1}^n f_i(b_i) \geq \sum_{i=1}^n f_i(b_{\pi(i)}) \geq \sum_{i=1}^n f_i(b_{n-i+1})$$

with $\big(f_{i+1}(x) - f_i(x)\big)$ nondecreasing for all $1 \leq i < n$.

Dually: $\prod_{i=1}^n (a_i + b_i) \leq \prod_{i=1}^n (a_i + b_{\pi(i)}) \leq \prod_{i=1}^n (a_i + b_{n-i+1}) \quad \text{for } a_i, b_i \geq 0.$

Cheat Sheets

[Optimization inequalities cheat sheet]

2. *f is convex*

a. $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$ for all $\lambda \in [0,1]$.

b. $f(x) \le f(y) + \langle \nabla f(x), x - y \rangle$

c. $0 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle$

d. $f(\mathbb{E}X) \le \mathbb{E}[f(X)]$ where X is a random variable (Jensen's inequality).

e. $x = \mathrm{prox}_{\gamma f}(x) + \gamma \mathrm{prox}_{f^*/\gamma}(x/\gamma)$, where f^* is the Fenchel conjugate and $\mathrm{prox}_{\gamma f}(x)$ is the proximal operator of γf. This identity is sometimes referred to as Moreau's decomposition

5. *f is both L-smooth and μ-strongly convex.*

a. $\frac{\mu L}{\mu + L}\|x-y\|^2 + \frac{1}{\mu+L}\|\nabla f(x) - \nabla f(y)\|^2 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle$

b. $\mu \preceq \nabla^2 f(x) \preceq L$ (assuming f is twice differentiable)

c. $f(x) \le f(y) + \langle \nabla f(x), x - y \rangle - \frac{\mu}{2}\|x-y\|^2 - \frac{1}{2(L-\mu)}\|\nabla f(x) - \nabla f(y) - \mu(x-y)\|^2$

Cheat Sheets

[The Matrix Cookbook]

2.4 Derivatives of Matrices, Vectors and Scalar Forms

2.4.1 First Order

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{69}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{70}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{71}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{72}$$

$$\frac{\partial \mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} \tag{73}$$

$$\frac{\partial (\mathbf{X}\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{im}(\mathbf{A})_{nj} = (\mathbf{J}^{mn}\mathbf{A})_{ij} \tag{74}$$

$$\frac{\partial (\mathbf{X}^T\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{in}(\mathbf{A})_{mj} = (\mathbf{J}^{nm}\mathbf{A})_{ij} \tag{75}$$

2.4.2 Second Order

$$\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl}X_{mn} = 2\sum_{kl} X_{kl} \tag{76}$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \tag{77}$$

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{B}\mathbf{x}+\mathbf{b}) \tag{78}$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})_{kl}}{\partial X_{ij}} = \delta_{lj}(\mathbf{X}^T\mathbf{B})_{ki} + \delta_{kj}(\mathbf{B}\mathbf{X})_{il} \tag{79}$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})}{\partial X_{ij}} = \mathbf{X}^T\mathbf{B}\mathbf{J}^{ij} + \mathbf{J}^{ji}\mathbf{B}\mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{80}$$

See Sec 9.7 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

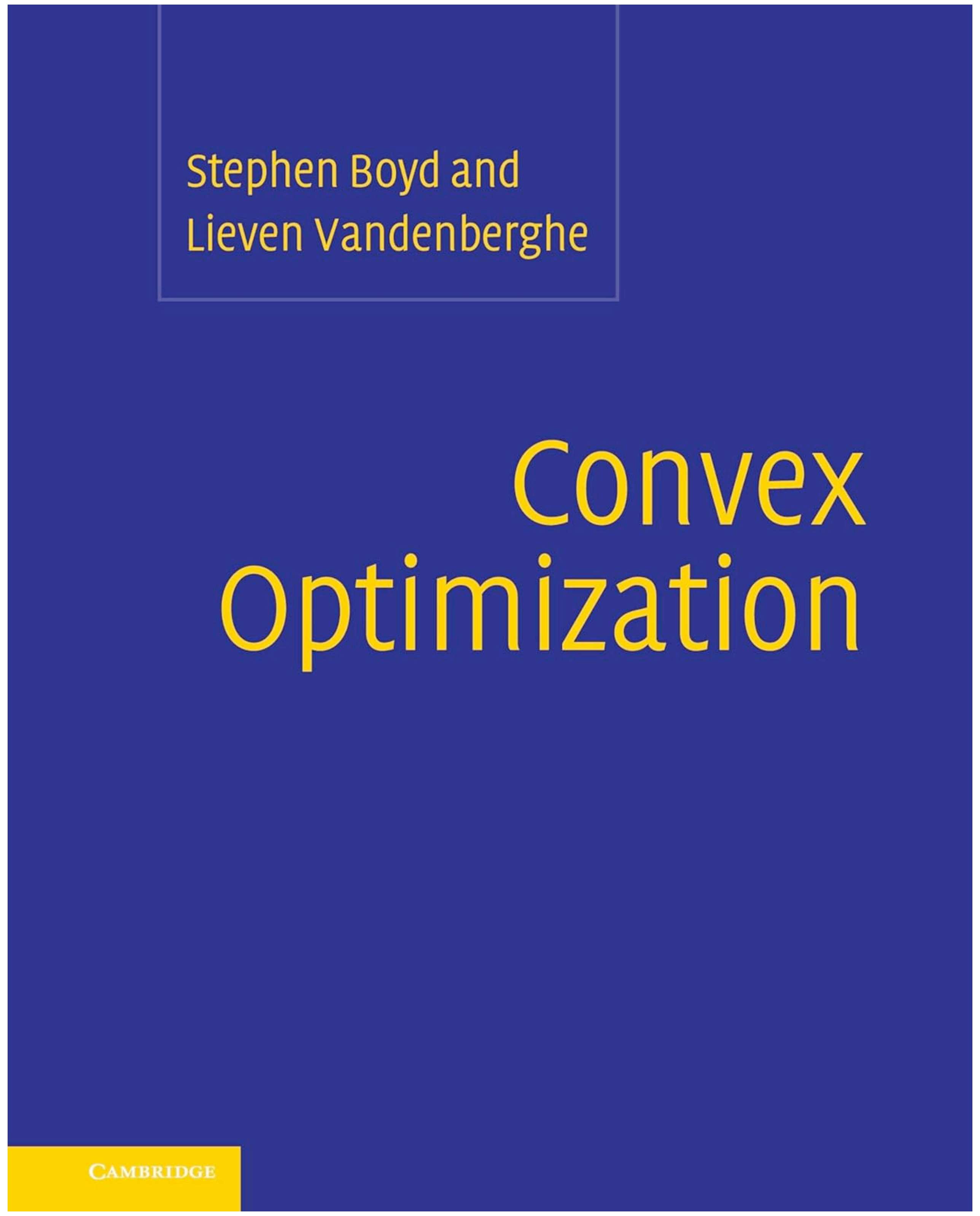
$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x} \tag{81}$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T\mathbf{X}\mathbf{b}\mathbf{c}^T + \mathbf{D}\mathbf{X}\mathbf{c}\mathbf{b}^T \tag{82}$$

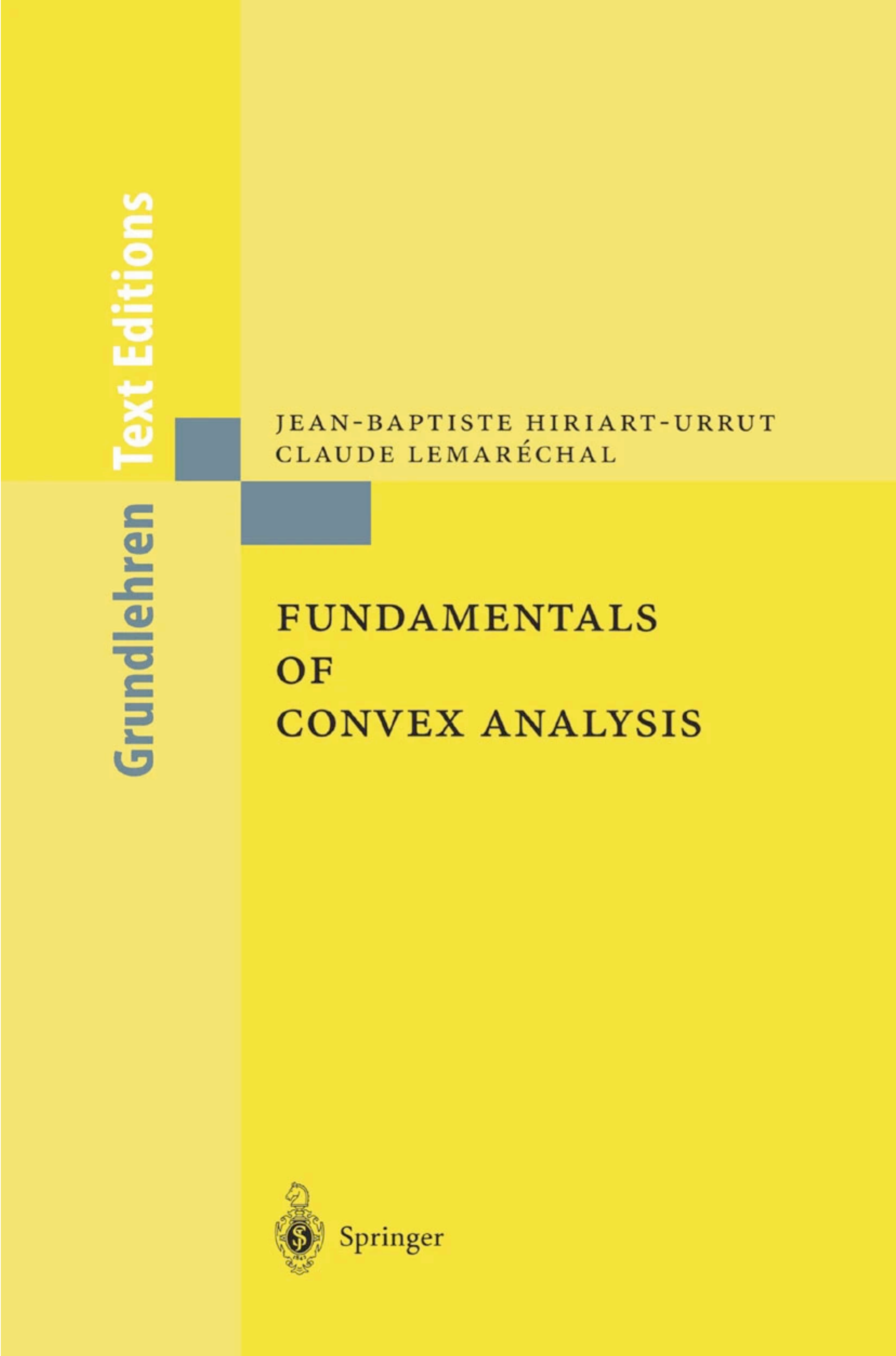
$$\frac{\partial}{\partial \mathbf{X}}(\mathbf{X}\mathbf{b}+\mathbf{c})^T\mathbf{D}(\mathbf{X}\mathbf{b}+\mathbf{c}) = (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T \tag{83}$$

Books - Optimization

Convex Optimization - Comprehensive, basic introduction to convex optimization. Goal is to teach how to recognize, formulate, and solve convex optimization problems.

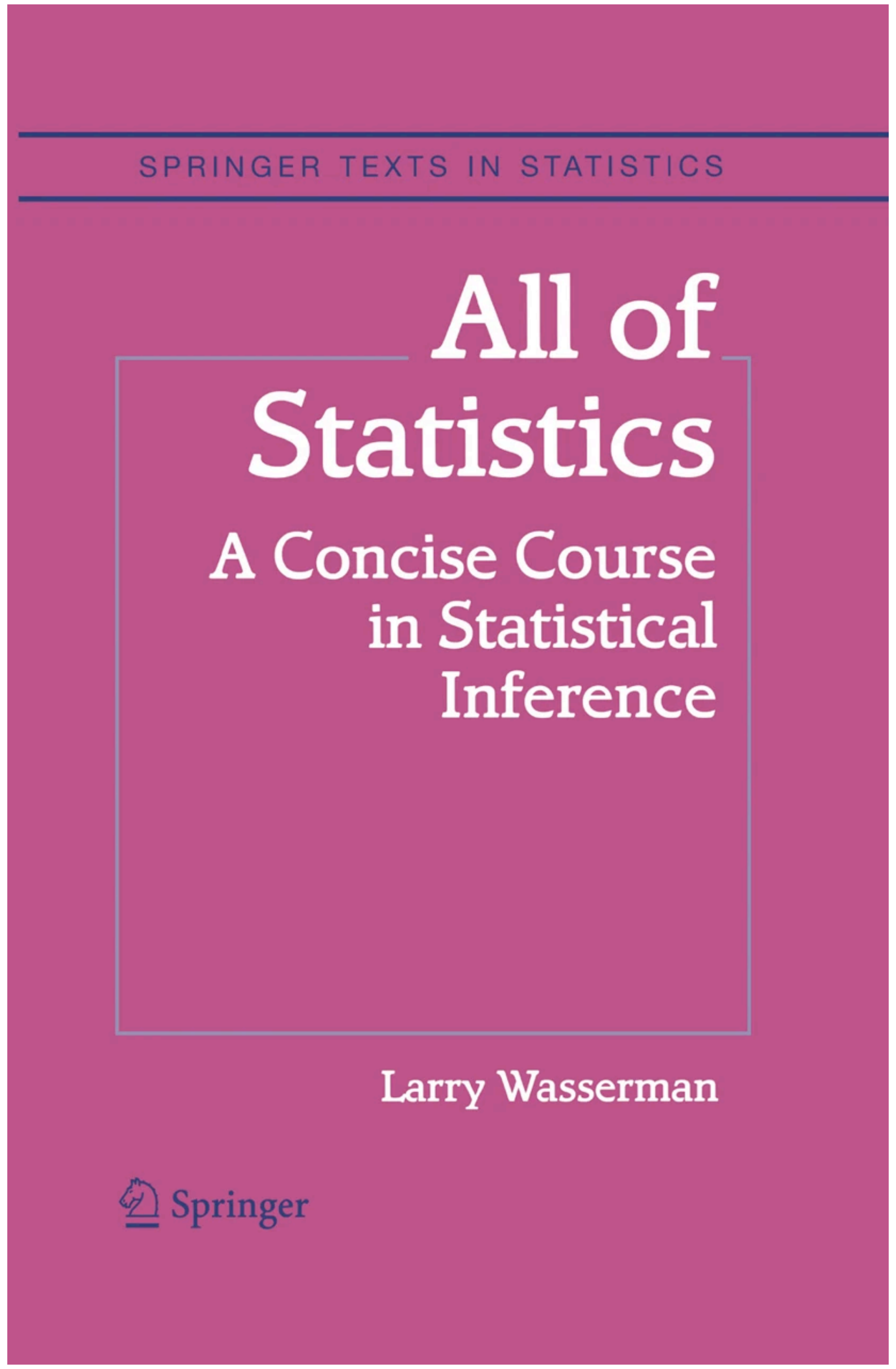


Fundamentals of Convex Analysis - deep but not too advanced introduction to convex analysis

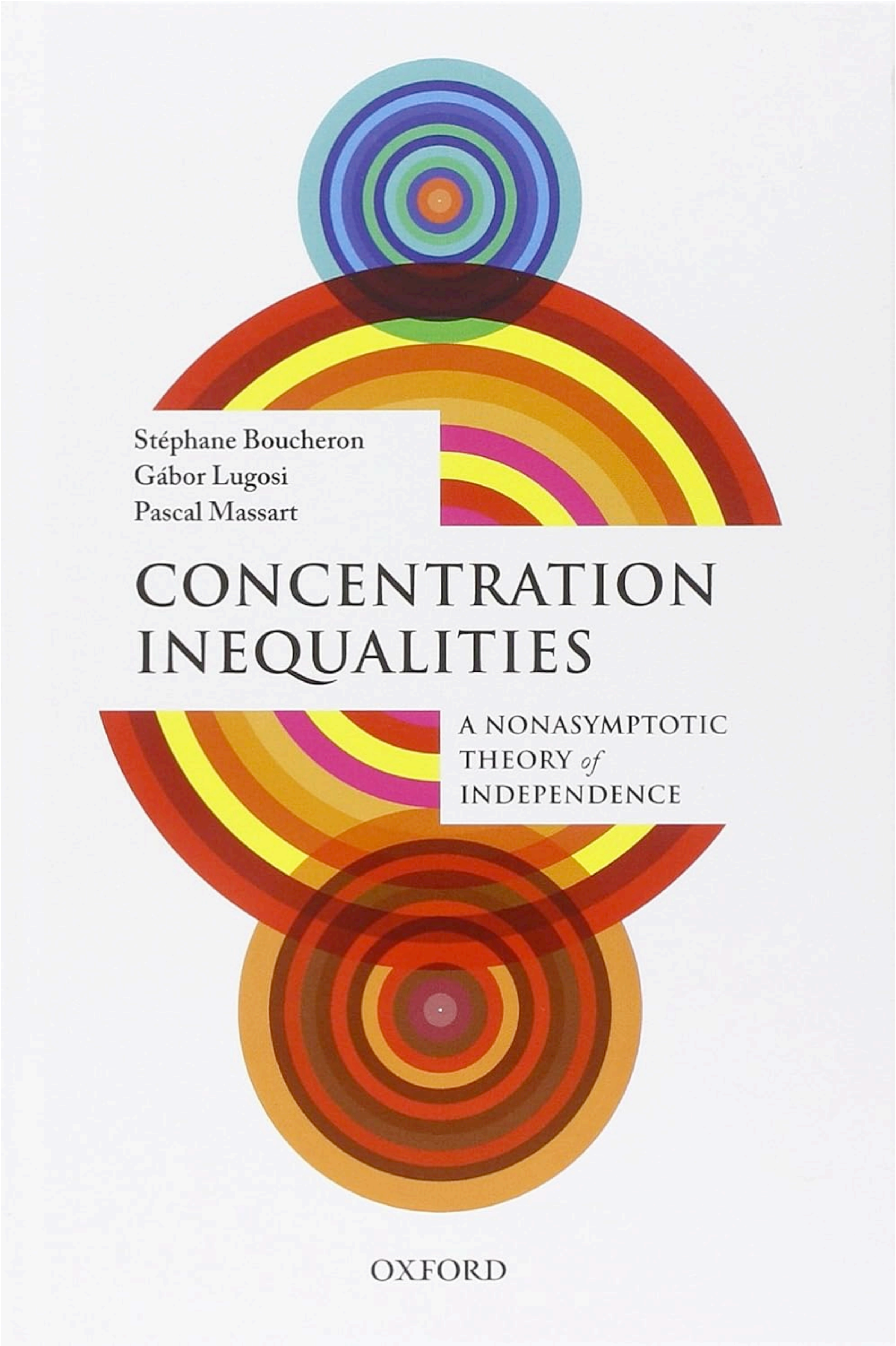


Books - Statistics

All of Statistics - quickly get back to up speed with fundamentals, and more

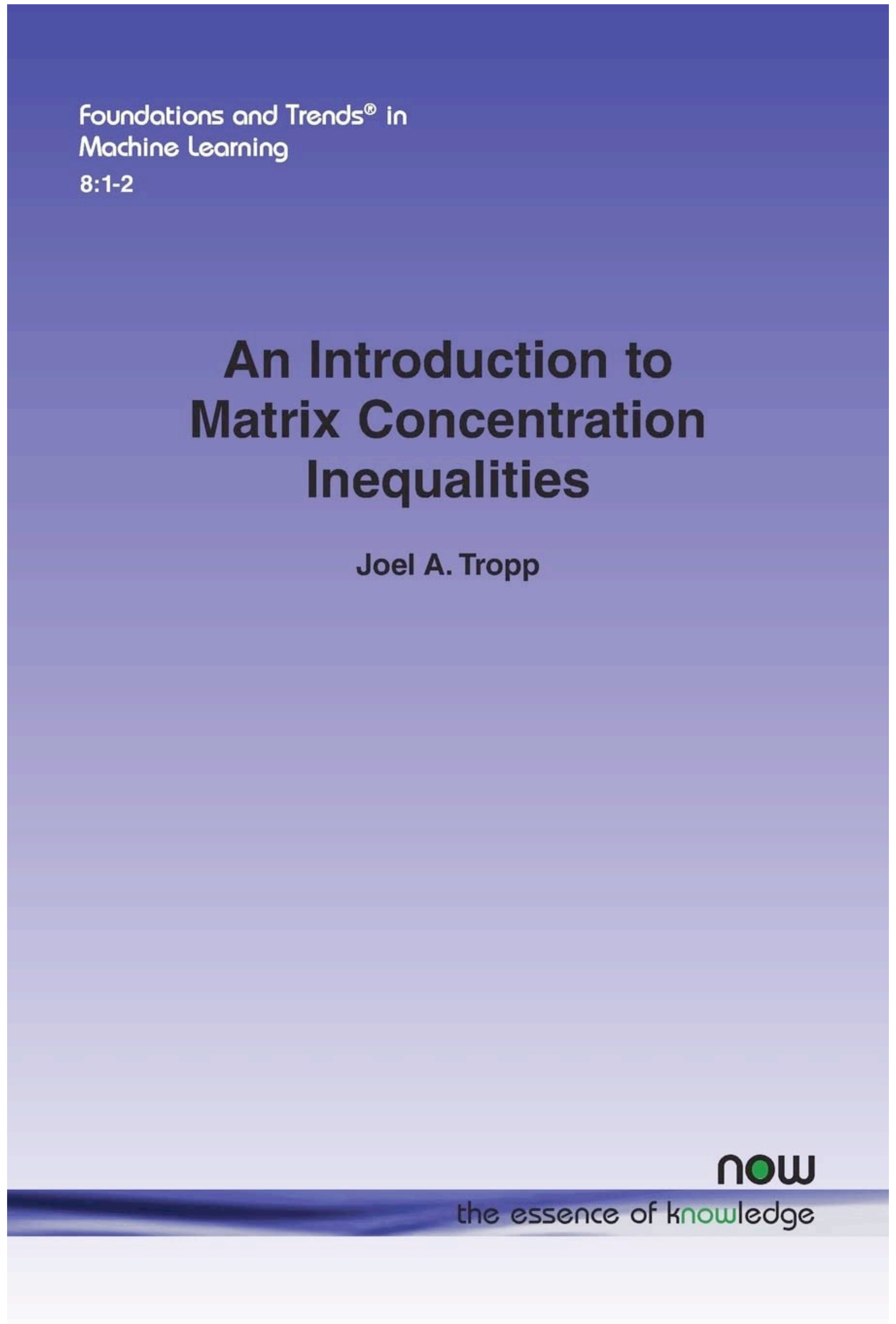


Concentration Inequalities - covers both basic and quite advanced results



Monograph - Random Matrices

An Introduction to Matrix Concentration Inequalities
- generalizations of standard concentration inequalities to matrices, and more



Tools/Software - General Math

- Mathematica and Maple - very useful for symbolic manipulation (avoid mistakes!)
- SageMath (free alternative)

Tools/Software: Optimization

- CVX
  - Using MATLAB: https://cvxr.com/cvx/doc/quickstart.html
  - Using Python: https://www.cvxpy.org

Techniques

- Working in discrete time and stuck?
  - Switch to continuous time. Sometimes things become simpler
- Working in continuous time? Lacking intuition?
  - Try discretizing to run experiments and build intuition

Techniques - Minimax Lower Bounds

- Try to prove information-theoretic lower bound:
  - Bin Yu's paper “Assouad, Fano, and Le Cam”
  - Introduction to Nonparametric Estimation (book, by Tsybakov)

Techniques - Languages

- Express things in a different language
  - For convex analysis, express things in language of Bregman divergences

Techniques - Experts

- Put in fair effort, and then try:
    - MathOverflow
    - LLM Chatbot
    - Ask advisor about experts in department, in university, or their contacts outside university

General Advice when Trying to Prove an Inequality

those key technical lemmas in papers

- "Behind every great theorem likes a great inequality" – paraphrasing Kolmogorov
- **Problem:** Proving inequality is hard work. Unclear if inequality is true. How much effort to spend?
- **Low-Effort Strategy 1:** Get visual intuition - Fix all but one variable and plot inequality (Desmos).
- **Low-Effort Strategy 2:** Same as above, but do 3D plot (Desmos can do 3D plots)
- **Medium-Effort Strategy:** Initially, play with concrete examples to try to show inequality *doesn't* hold. The more examples (and the more clever the examples) that do not violate inequality, the more motivation to spend time proving inequality *does* hold.

General Advice when Trying to Prove an Inequality

- **Problem:** I tried those other strategies and am still stuck. How I try to prove the inequality?
- **"Taylor'ing" Strategy (for functions of continuous variables):**
    - To show $f(x) \geq g(x)$:
        - Use first-order Taylor approximation to lower bound $f(x) - g(x)$ by $h(x)$.
        - Try to lower bound $h(x)$ by 0.
        - If no success, try again with second-order Taylor expansion. If again no success, third-order, fourth-order, …

General Advice when Trying to Prove an Inequality

- **Problem:** I tried everything and am *still* stuck. Also, I can't seem to prove the inequality is true. What now?
- **"Opposite" Strategy:** Try to prove that the inequality is NOT true
- Sometimes you end up with a proof that inequality is true AND proof that inequality is not true
  - Now begins the fun: figure out which proof is wrong :-)