1. The core concept of convex optimization is to prove any directions from feasible solution of a convex function is larger than zero, it should be emphasized in every method to prove it's the descent direction, on the contrary, it's the same as ascend direction when it's concave.
2. Recommend emphasizing optimizers in every problem, e.x. inner product optimizers in complete inner-product spaces.
3. Gradient descent and its families are usually unconstrained problems, they don't apply on convex sets unless they are proven locally continuous and differentiable (Lipschitz continuity), Lipschitz continuity should be proven if used, otherwise, it should be briefly explained why it applies.
4. Following the previous statement, recommend adding constrained and unconstrained problems since they have different approach.
5. Recommend briefly explain constrain function of each problem, especially when it's hyper-plane.
6. Projection Theorem is feasible only if the target space is Hilbert or complete Euclidean space.

   $$\textit{The vector space } X \subseteq H \textit{ or giving proper Euclidean properties.}$$

   $$d: \mathbb{R}^D \rightarrow \mathbb{R}, d_n(x, y) = \sqrt[n]{(x-y)^n}.$$

7. Duality Theorem is feasible without Karush–Kuhn–Tucker conditions, recommend adding additional section to explain KKT conditions.
8. Instead of using equations to express Prime-Dual algorithm, using matrix expression is better for implementation and understanding.
9. Highly recommend adding some basic concepts of Information Theory.
10. Recommend adding KL-divergence and its properties when it's convex sets.
11. Readers are suggested to have basic concepts of stochastic processes to understand methods involved with randomness.
12. Besides stochastic processes, in order to introduce Monte-Carlo method, it's recommended to start with random walks.
13. Markov process and Gibbs sampler are recommended to understand MCMC.
14. Recommend adding Dynamic programming method since it's a simple version of MCMC.
15. Every method that involves with conditional probabilities should be explained more clearly for sampling relationship between distributions.

16. Any commonly used notations should be avoided when defining, e.x. $\boldsymbol{J}$ as Jacobian, $\boldsymbol{I}$ as identity matrix and $\boldsymbol{Q}$ as Q-function.
17. Equation (3.83) should be

$$p_{X,Y|\Theta}(x,y|\theta) = p_{X|Y,\Theta}(x|y,\theta)p_{Y|\Theta}(y|\theta), then\ p_{X|\Theta}(x|\theta)$$

$$= \sum_{y \in Y} p_{X,Y|\Theta}(x,y|\theta).$$

18. Every problem that involves with probabilities and stochastic processes can be discussed with additional chapters since they have comparably different approach and they are not limited to convex optimization.
19. $\sigma - algebra$ should be defined when defining additional measurements in vector spaces.
20. Several notations misused, such as P in probability represents measurements of certain events, if it represents density, it should be f, or p for mass.

$$P[X \leq x],$$
$$the\ probability\ of\ an\ event\ that\ contains\ outcomes\ in\ X\ smaller\ than\ x.$$
$$f_X(x)\ or\ p_X(x),$$
$$the\ probability\ measure\ of\ outcome\ x\ in\ distribution\ of\ X.$$

21. Following the tutorial of Maximum-likelihood Estimation, it seems like it's following Weak rule of Big Numbers, the explanation can be clearer on convergence such as distribution, probability or expectation.