1. p. 15, Eqn. 2.10-2.11. Here and in Exercise 2.2 and 2.3, it would be helpful to reiterate that data are standardized <z,z>/N = 1.
2. p. 19, 2nd sentence of Section 2.6: Every lasso problem has a solution with at most N nonzero entries but not every minimizer has to have N or less nonzero entries. Would be nice to see the original reference Osbourne et al. (2000) in the book.
3. p. 89, Exercise 4.12: \hat g_2 -> \hat g_j.
4. p. 102: $\alpha\in(0,1)$ in Armijo rule.
5. p. 107: “+” in (5.31)
6. Chap 5.3.2: more detailed on how to do the backtracking line search would be nice and perhaps reference to the Beck and Teboulle (2009) chapter?
7. p. 121, line after (5.62): nxd -> dxm, nxd -> dxn
8. p. 137, Ex. 5.12 (a) (b)
9. p. 273, \Phi instead of \Phi^T in (10.4)
10. p. 277, line -1: “Boolean vectors $u\in\{-1,0,1\}^p$”.
11. p. 278, line 2: seems more natural to use log(2e…) instead of log(e^2…).
12. p. 281, 10.4.1, C(S) is not a convex cone, just a cone.
13. p. 282, equation (10.22): Add j\not= j’
14. p. 284, two lines after (10.24): extra | before double sum on the RHS.
15. p. 286, Ex. 10.1 (a): when d = N. (b) would be helpful to state that the upper bound is not tight / what the best constants and the largest interval for t are.
16. p. 286, Ex. 10.2 (a): ||u||_2^2 = 1.
17. p. 287, Ex. 10.6 (b): X_S should be X (otherwise not conformable)
18. p. 286, Ex. 10.3(c): Replace the bound M = (1/ \epsilon)^{cp} with M = (c / \epsilon)^p
19. p. 286, Ex. 10.3(a): The claim is not true. Indeed, for any numbers $\epsilon > 0$ and $\alpha > 0$, there exists a minimal ϵ-covering set of a compact set $\mathcal{A}$ that is not an α-packing set. To see this, choose N so that $\frac{\epsilon}{N} \leq \alpha$ and consider the interval $\mathcal{A} = [0, \left(3+ \frac{1}{N} \right) \epsilon]$. The set $\mathcal{U} = \left\{ \epsilon, \left(1 + \frac{1}{N} \right) \epsilon, \left(3 + \frac{1}{N} \right) \epsilon \right\}$ is a minimal ϵ-covering of $\mathcal{A}$, but not an $\frac{\epsilon}{N}$-packing set (and hence not an α-packing set).
20. p. 286, Ex. 10.3(b): Every maximal ϵ-packing set is an ϵ- covering set.
21. p. 295, line -5: no need for “stochastically dominated by”
22. p. 300, period at end of (11.25a).
23. p. 301, proof of (11.25b): why not use (11.23) instead of (11.22)? And 3 instead of 12 as multiplicative constant?
24. p. 303, (11.31): Has matrix infinity norm been defined before?
25. p. 305, (11.32): “-” missing?
26. p. 308: “-” RHS in (11.35)?
27. p. 309, last two displays: Z_S instead of U_S.
28. p. 309, line -8/9: “choice of \lambda_N guarantees…” some extra hypothesis seems needed as \lamda_N in chosen as a function also of \sigma and K_clm which could be very small.