Optimization and Computational Linear Algebra for Data Science
Hints for the review exercises

hints. please only look at the hints if you have spent a reasonable time thinking about the problems!

1 2019 review exercises

1. Use the fact that $\|Ax\|^2 = x^\top A^\top Ax$ and then use the SVD decomposition of A to rewrite $A^\top A$.
2. Use the SVD of A.
3. (a) True (b) False (c) False (eigenvalues can be negative but singular values can not. The singular values of a symmetric matrix are the absolute value of its eigenvalues).
4. Use the definitions of kernel and image.
5. Use the SVD decomposition of A to compute $(A^\top A)^{-1}A^\top$ and see that it corresponds to the definition of †.
6. Convex, convex, not convex.
7. Convex, not convex, convex.
8. Express the columns of B using the left-singular vectors of the matrix A whose rows are the a_i.
9. False. False. True.
10. Show that $(x+y)/2$ is a global minimizer of f.
11. Normalizing the dataset is useless for ordinary least-squares, but can be useful for Lasso.
12. Compute gradient and Hessian.
13. Use Lagrange multipliers.
14. If the columns of A are linearly dependent, then f will be L-smooth but not strongly convex, hence the speed of gradient descent will be $O(1/t)$. If the columns of A are linearly independent then you can show that $f(x)$ is μ-strongly convex and L-smooth, for some $\mu, L > 0$. Hence the error of gradient descent will be $O(e^{-\rho t})$ after t steps, for some constant $\rho > 0$.
15. 1 step.
16. See lecture notes.

2 2018 review exercises

1. Show that for all $x \in \mathbb{R}^n$, $ABx = BAx$. (You can decompose such x in the given basis)
2. (a) See homework 10. (b) Use (a).
3. Use the definition of eigenvectors/eigenvalues.
4. Using the spectral Theorem there exists an orthonormal basis $(v_1, \dots, v_n)$ of $\mathbb{R}^n$ consisting of eigenvectors of A. Decompose x in such a basis and compute Ax.
5. If $f : \mathbb{R}^2 \to \mathbb{R}$ is convex, and if (α^*, β^*) is a minimizer of f then $\nabla f(\alpha^*, \beta^*) = 0$.
6. Use the definition of $\|x\|_\infty$ and $\|x\|$.
7. By the spectral theorem, you can decompose x in an orthonormal basis of $\mathbb{R}^n$ made of eigenvectors of A.
8. Many possible ways to do this. (a) Show that $\mathrm{Ker}(A^\mathsf{T}) = \mathrm{Ker}(AA^\mathsf{T})$, and then use the rank-nullity theorem and the fact that $\mathrm{rank}(A) = \mathrm{rank}(A^\mathsf{T})$. (b) Compute AA^T using the SVD of A: $A = U\Sigma V^\mathsf{T}$.
9. (a) Use Lagrange multipliers. (b) The set of solution of $Ax = b$ is $A^+b + \mathrm{Ker}(A)$. The result follow from the same arguments than problem 1 of homework 10.
10. False.
11. Show that $\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2$.
12. Show that $\sum_{i=1}^n \langle x, u_i\rangle^2 = \|x\|^2$.
13. Compute AA^T.
14. Show that if λ is an eigenvalue of A associated with the eigenvector u if and only if Qu is an eigenvector of B with eigenvalue λ.
15. Justify that $x = \sum_{i=1}^m \langle x, v_i\rangle v_i$. Then expand $\| \sum \langle x, v_i\rangle v_i\|^2$ and make simplifications.
16. Use the SVD of A.
17. (a) See Homework 3. (b) Let $V \in \mathbb{R}^{n\times n}$ be the matrix whose columns are $v_1, \dots, v_n$. Show that $\mathrm{Tr}(V^T Av) = \sum_{i=1}^n v_i^\mathsf{T} A v_i$. Then use (a). (c) Use the spectral theorem and (a).
18. Use Problem 1.b from homework 7.
19. Expand the right-hand side.
20. (a). $A^2 = 0$. (b) Take for instance
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
21. (a) convex, (b) not convex (c) not convex (d) convex. One can verify these points by computing the Hessian.
22. (a) ... (b) There is a unique global minima.
23. V is of dimension 2, hence $\dim(V^\perp) = 4 - 2 = 2$. $v_1 = (1, -1, 0, 0)$ and $v_2 = (0, 0, 1, -1)$ work.

24. (a) Yes. (b) The derivative of a sum is equal to the sum of the derivaties, and the derivatives of λp (for some $\in \mathbb{R}$) is equal to $\lambda p'$. (c) $\mathrm{Ker}(\mathcal{D})$ is the set of polynomials p that are constant (i.e. there exists $a \in \mathbb{R}$ such that $p(x) = a$ for all $x \in \mathbb{R}$). (d) $\mathrm{Im}(\mathcal{D}) = \mathcal{P}_{d-1}$. (e) (i) check the usual conditions (ii) For polynomial of degree $\leq d$, Taylor formula of order d is exact:

$$T_s(p)(x) = p(x+s) = \sum_{k=0}^{d} \frac{p^{(k)}(x)}{k!} s^k = \sum_{k=0}^{d} \frac{\mathcal{D}^k(p)(x)}{k!} s^k.$$

(iii) The matrix has 0 below the diagonal and for $j \geq i$, $M_{i,j} = \binom{j-1}{i-1}$.

25. (a) Let B be a rank 1 matrix. One can therefore write $B = uv^T$ for some $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$.

$$\|A - B\|_F^2 = \|A\|_F^2 - 2u^T A v + \|u\|^2 \|v\|^2.$$

Now, $u^T A v \leq \|u\| \|v\| \sigma_1$. Hence, writing $r = \|u\| \|v\|$

$$\|A - B\|_F^2 \geq \sum_{i=1}^{\min(n,m)} \sigma_i^2 - 2\sigma_1 r + r^2 = \sum_{i=2}^{\min(n,m)} \sigma_i^2 + (\sigma_1 - r)^2 = \|A - A'\|_F^2 + (\sigma_1 - r)^2$$

(b) Let $B = uv^T$ be a rank 1 matrix. Let v_1, v_2 be the first two right-singular vectors of A. $Span(v)^\perp$ has dimension $n-1$, hence one can find a vector of unit norm z in $Span(v)^\perp \cap \mathrm{Span}(v_1, v_2)$. We write $z = \alpha_1 v_1 + \alpha_2 v_2$. Since $\|z\| = 1$ and v_1, v_2 orthogonal, we have $\alpha_1^2 + \alpha_2^2 = 1$. By definition of the spectral norm

$$\|A - B\|_{\mathrm{Sp}} \geq \|(A - B)z\| = \|Az\| = \sqrt{\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2} \geq \sigma_2 = \|A - A'\|_{\mathrm{Sp}}.$$

26. xx^T is rank 1 and has two distinct eigenvalues 0 and 1. Hence H has two distinct eigenvalues -1 and 1.

27. The vector $(1, 1, \ldots, 1)$ is an eigenvector associated with the eigenvalue d. By contradiction let x be an eigenvector associated with the eigenvalue $\lambda > d$. Let i such that $|x_i| = \|x\|_\infty > 0$. Then

$$|x_i|\lambda = |\sum_{j=1}^{n} G_{i,j} x_j| \leq \sum_{j=1}^{n} G_{i,j} |x_j| \leq |x_i| \sum_{j=1}^{n} G_{i,j} = d|x_i|.$$

We get a contradiction.

28. Use the matrix product formula.

29. (a) convex but not subspace (b) not convex (c) subspace (hence convex)

30. (a) convex but not strictly convex (b) convex but not strictly convex (c) convex but not strictly convex (d) not convex (e) not convex

31. Cauchy-Schwarz.

32. Apply the spectral theorem to A.

3 2018 final

1. (a) Show that $w \overset{\text{def}}{=} u - v \neq 0$ belongs to $\mathrm{Ker}(A)$ and then that $x = u + tw$ is a solution to $Ax = b$ for all $t \in \mathbb{R}$. (b) Use the w defined in (a).

2. (a). Prove that $\text{rank}(A^T A) = n$. Then use the fact that $\text{rank}(A^T A) \leq \text{rank}(A) \leq \min(n, m)$. (b) Show that $\text{rank}(A) = n$, then use the rank-nullity theorem to get $\text{dimKer}(A) = 0$. (c) Use the fact that for any $v \in \mathbb{R}^n$, $\|Av\|^2 = v^\mathsf{T} A^T A v$.

3. (a) $\text{rank}(U) = n$. This comes from the fact that $\text{rank}(U) \leq n$ (because U is $n \times m$) and $\text{rank}(U) \geq \text{rank}(U^T U) \leq \text{rank}(\text{Id}_n) = n$. (b) Use the rank-nullity theorem. (c) Expand and simplify $\|y - Ux\|^2$. (d) Write $f(x) = \|y - Ux\|^2$ and solve $\nabla f(x) = 0$.

4. (a) Convex set ($\text{Ker}(A)$) (b) Convex set (c) Not always convex (take for instance for $A = \text{Id}$) (d) Not convex. (e) Convex.

5. (a)(e)(d)(f)

6. (a) f is convex (compute its Hessian). (b) h is not convex (compute the Hessian) (c) Solve the equations $->$ Global minimum (d) Solve the equations $->$ Saddle point

7. (a) By contradiction if for all j we have $|v_i^T u_1| < \frac{1}{\sqrt{n}}$ then we get (since $v_1, \dots, v_n$ orthonormal basis)

$$\|u_1\|^2 = \sum_{i=1}^{n} (v_i^T u_1)^2 < 1,$$

which is a contradiction. (b) Take $u_1 = (1, 0)$, $u_2 = (0, 1)$, $v_1 = (1/\sqrt{2}, 1/\sqrt{2})$, $v_2 = (1/\sqrt{2}, -1/\sqrt{2})$.