Alberto Quaini

Exercise 6.1

Let $f(w) := -e^{-w^T x}$, $G(w) := w^T A w - w^T A y - w^T x$, and $H(w) = y^T w - w^T x$. Then the problem can be written in the following standard form:

$$
\begin{aligned}
\text{minimize} \quad & f(w) \\
\text{subject to} \quad & G(w) \leq a \\
& H(w) = b.
\end{aligned}
$$

Exercise 6.6

The gradient is $Df(x,y) = (6xy + 4y^2 + y, 3x^2 + 8xy + x)$ and the hessian is

$$
D^2 f(x,y) = \begin{bmatrix} 6y & 6x + 8y + 1 \\ 6x + 8y + 1 & 8x \end{bmatrix}.
$$

The first order conditions read $Df(x,y) = (0,0)$ and yield the following critical points: $A = (-1/3, 0)$, $B = (-1/9, -1/12)$, $C = (0,0)$ and $D = (0, -1/4)$. The eigenvalues of the hessian for A are approximately 0.3 and -3, thus A is a saddle point. The eigenvalues of the hessian for B are approximately -0.3 and -1.1, thus B is a local maximizer. The eigenvalues of the hessian for C are approximately 1 and -1.1, thus C is a saddle point. The eigenvalues of the hessian for D are approximately -2 and 0.5, thus D is a saddle point.

Exercise 6.7

(i)

Notice that $Q^T = (A^T + A)^T = A^T + A = A + A^T = Q$. Also, $x^T A x = \sum_{i=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ji} x_i x_j = x^T A^T x$. Therefore $x^T Q x = 2x^T A x$ and (6.17) is equivalent to

$$
f(x) = x^T Q x / 2 - b^T x + c.
$$

(ii)

The first order necessary conditions for a minimizer imply $Q^T x^* = b$, since $f'(x) = Q^T x - b$.

(iii)

If Q is positive definite, then $f''(x) > 0$ for any x. Also, Q is invertible and by (6.19) we have that $x^* = Q^{-1} b$ is such that $f'(x^*) = 0$. Then by the second order sufficient condition, x^* is the unique minimizer of f. Now assume x^* is the unique minimizer of f. Then by the second order necessary condition, Q is positive semi-definite. Also, x^* is a solution to $Q^T x^* = b$. If Q has at least one zero eigenvalue, then x^* is not unique. Therefore Q must be positive definite.

Exercise 6.11

Notice that $f'(x) = 2ax + b$, $f''(x) = 2a$, and that the first Newton's Method iteration is $x_1 = x_0 - f'(x_0)/f''(x_0)$. Notice that

$$f'(x_1) = 2a(x_0 - (2ax_0 + b)/2a) + b = 0$$

and

$$f''(x_0) = 2a > 0.$$

Therefore, x_1 is a local minimizer. Since f is quadratic, it is the unique minimizer.

0.1 Exercise 7.1

Take $x, y \in \text{conv}(S)$. Then $x = \sum_{i=1}^{k_x} \lambda_i^x s_i$ where s_i are elements of S, $k_x \in \mathbb{N}$ and λ_i^x are nonnegative and sum to 1. Do the same for y and set $k = \max\{k_x, k_y\}$. Also, let $\lambda \in [0, 1]$. Then

$$\lambda x + (1 - \lambda)y = \sum_{i=1}^{k} (\lambda\lambda_i^x + (1 - \lambda)\lambda_i^y)s_i$$

where $(\lambda\lambda_i^x + (1 - \lambda)\lambda_i^y)$ are nonnegative and

$$\sum_i \lambda\lambda_i^x + (1 - \lambda)\lambda_i^y = \lambda \sum_i \lambda_i^x + (1 - \lambda) \sum_i \lambda_i^y = 1.$$

Thus $\lambda x + (1 - \lambda y) \in S$ and S is convex.

Exercise 7.2

(i) Let $P = \{x \in V \ : < a, x >= b\}$ for some $a \in V$, $a \neq 0$ and some real b. Let $x, y \in P$ and $0 \leq \lambda \leq 1$. Then

$$< a, \lambda x + (1 - \lambda)y >= \lambda < a, x > +(1 - \lambda) < a, y >= \lambda b + (1 - \lambda)b = b.$$

Thus P is convex.

(ii) Let $H = \{x \in \mathbb{R}^n \ : < a, x >\leq b\}$ where again $a \in V$, $a \neq 0$ and some real b. Let $x, y \in H$ and $0 \leq \lambda \leq 1$. Then

$$< a, \lambda x + (1 - \lambda)y >= \lambda < a, x > +(1 - \lambda) < a, y >\leq \lambda b + (1 - \lambda)b = b.$$

Thus H is convex.

$$< a, \lambda x + (1 - \lambda)y >= \lambda < a, x > +(1 - \lambda) < a, y >\leq \lambda b + (1 - \lambda)b = b.$$

Exercise 7.4

(i) Note that

$$||x-y||^2 = ||(x-p)+(p-y)||^2 =$$
$$< (x-p)+(p-y),(x-p)+(p-y) > = ||x-p||^2 + ||p-y||^2 + 2 < x-p, p-y > .$$

(ii) Take an arbitrary $y \neq p$. Then $||p-y||^2 > 0$. Suppose $< x-p, p-y > \geq 0$. Then it is clear by (i) that $||x-y|| > ||x-p||$.

(iii) Let $z = \lambda y + (1-\lambda)p$, where $0 \leq \lambda \leq 1$. Then by (i) where we use z instead of y we get

$$||x-z||^2 = ||x-p||^2 + ||\lambda y - \lambda p||^2 + < x-p, \lambda p - \lambda y > =$$
$$||x-p||^2 + 2\lambda < x-p, p-y > + \lambda^2 ||y-p||^2.$$

(iv) In (7.15), put $\lambda = 1$, so $z = y$. Then we know that

$$0 \leq ||x-y||^2 - ||x-p||^2 = 2\lambda < x-p, p-y > + \lambda^2 ||y-p||^2.$$

Dividing by λ you get $0 \leq 2\lambda < x-p, p-y > + \lambda^2 ||y-p||^2$. Take $y = p$, then $0 \leq \lambda < x-p, p-y >$.

The if statment of the theorem follows by (iv). The only if statment of the theore follows by (ii).

Exercise 7.8

Let $x, y \in \mathbb{R}^n$, with $x \neq y$, and $\lambda \in [0,1]$. Then

$$g(\lambda x + (1-\lambda)y) = f(\lambda Ax + (1-\lambda)Ay + b) = f(\lambda(Ax+b) + (1-\lambda)(Ay+b))$$
$$\leq \lambda f(Ax+b) + (1-\lambda) f(Ay+b) = \lambda g(x) + (1-\lambda) g(y)$$

shows that g is convex.

Exercise 7.12

(i)

Take $X, Y \in PD_n(\mathbb{R})$ and $\lambda \in [0,1]$. Then for every $v \in \mathbb{R}^n$ we have that

$$v^T(\lambda X + (1-\lambda)Y)v = \lambda(v^T X v) + (1-\lambda)(v^T Y v) > 0,$$

because X and Y are positive definite.

(ii)

(a) Take $t_1, t_2 \in \mathbb{R}$ and $\lambda \in [0,1]$. On the one hand,

$$\lambda g(t_1) + (1-\lambda) g(t_2) = \lambda f(t_1 A + (1-t_1)B) + (1-\lambda) f(t_2 A + (1-t_2)B).$$

On the other,

$$\begin{aligned} g(\lambda t_1 + (1-\lambda)t_2) &= f((\lambda t_1 + (1-\lambda)t_2)A + (1 - \lambda t_1 + (1-\lambda)t_2)B) \\ &= f(\lambda(t_1 A + (1-t_1)B) + (1-\lambda)(t_2 A + (1-t_2)B)). \end{aligned}$$

Since g is convex we get

$$f(\lambda X + (1-\lambda)Y) \leq \lambda f(X) + (1-\lambda)f(Y),$$

with $X = t_1 A + (1-t_1)B$ and $Y = t_2 A + (1-t_2)B$. Since the choice of t was arbitrary and this holds for any $A, B \in PD_n(\mathbb{R})$, we conclude that f is convex.

(b) By Proposition (4.5.7), we know that if A is posititve definite, then there exits a nonsingular matrix S such that $A = S^H S$. Then, $tA + (1-t)B = S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S$, and so

$$g(t) = -\log(\det(tA + (1-t)B)) = -\log(\det(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)).$$

By the fact that $\det(AB) = \det(A)\det(B)$ and the properties of logarithms, we obtain

$$\begin{aligned} -\log(\det(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)) &= -\log(\det(S^H)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})) - \log(\det(S)) \\ &= -\log(\det(S^H)\det(S)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})) \\ &= -\log(\det(A)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})). \end{aligned}$$

(c)

Since $A, B \in PD_n(\mathbb{R})$, then $B^{-1} \in PD_n(\mathbb{R})$ and $((S^H)^{-1}BS^{-1})^{-1} = SB^{-1}S^H$ is positive definite since

$$x^H SB^{-1}S^H x = (S^H x)^H B^{-1}(xS) > 0.$$

Therefore $(S^H)^{-1}BS^{-1}$ is positive definite. Now let $\{\lambda_i\}_i$ be the collection of eigenvalues of $((S^H)^{-1}BS^{-1})$ and $\{x_i\}_i$ the corresponding collection of eigenvectors. Then for every i:

$$(tI + (1-t)(S^H)^{-1}BS^{-1})x_i = tx_i + (1-t)\lambda_i x_i = (t + (1-t)\lambda_i)x_i.$$

Thus, $\{t + (1-t)\lambda_i\}_i$ are the eigenvalues of $(tI + (1-t)(S^H)^{-1}BS^{-1})$ corresponding to the $\{x_i\}_i$, and we can conclude that

$$\begin{aligned} -\log(\det(A)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})) &= -\log(\det(A)) - \log(\Pi_{i=1}^n(t + (1-t)\lambda_i)) \\ &= -\log(\det(A)) - \sum_{i=1}^n \log((t + (1-t)\lambda_i)). \end{aligned}$$

(d)

By using the expression of $g(t)$ in part (c) we can see that $g'(t)\sum_{i=1}^n(1-\lambda_i)/(t+(1-t)\lambda_i)$ and $g''(t) = \sum_{i=1}^n(1-\lambda_i)^2/(t+(1-t)\lambda_i)^2$, which is clearly nonnegative for all $t \in [0,1]$.

Exercise 7.13

Suppose $f(x) < M$ for all x for some real M and f is convex and not constant. Then, there exist $x, y \in \mathbb{R}^n$ such that $f(x) \neq f(y)$. But then the line between $(x, f(x))$ and $(y, f(y))$ intersects $f(\cdot) = M$. Since f must lie on or above this line, at some point it must cross $f(\cdot) = M$ as well, which is a contraddiction.

Exercise 7.20

Take $x, y \in \mathbb{R}^n$, with $x \neq y$, and $\lambda \in [0, 1]$. Since f is convex we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Since $-f$ is convex, the opposite hold. Therefore we must have $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$. Therefore f is affine.

Exercise 7.21

Let $x^* \in \mathbb{R}^n$ be a local minimizer of f. Then $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$, where $\mathcal{N}_r(x^*)$ is an open ball around x^* of radius $r > 0$. Since ϕ is monothonically increasing, $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of $\phi \circ f$. Now let x^* be a local minimizer of $\phi \circ f$. Then $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$, and since ϕ is monothonically increasing, this implies that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of f.