Problem Set 4

By: Bethany Bailey

6.6

Set the first derivative (with respect to x and y) equal to zero.

$\frac{df(x,y)}{dx} = 6xy + 4y^2 + y = 0$

$\frac{df(x,y)}{dy} = 3x^2 + 8xy + x = 0$

Solving this system and analyzing the eigenvalues of the Hessian at each solution, we get the following critical points:

$x = -1/3, y = 0$, eigvals of Hessian:0.3, -3, saddle point.

$x = -1/9, y = -1/12$, eigvals of Hessian: -0.3, -1.1, local maximizer.

$x = 0, y = -1/4$, eigvals of Hessian: 1, -1.1, saddle point.

$x = 0, y = 0$, eigvals of Hessian: -2, 0.5, saddle point.

6.7

(i)

$Q^T = (A + A^T)^T = A^T + A = A + A^T = Q$. Thus, Q is symmetric.

We also know that $x^T Qx = x^T(A + A^T)x = x^T Ax + x^T A^T x = 2x^T Ax$.

Thus, we know that (6.17) is equal to:

$f(x) = \frac{1}{2}x^T Qx - b^T x + c$

(ii)

To be a minimizer, the first derivative must equal 0. Taking the first order condition, we get

$$f'(x) = x^T Q - b$$

The solutions to this equation are found by setting it equal to 0, thus this equation can be reordered to $Q^T x = b$. Thus, any minimizer x^* of f is a solution of the equation $Q^t x^* = b$.

(iii)

Using the equation from (ii), we see that this has a unique solution only when Q^T is invertible, implying that Q is also invertible and has no 0 eigenvalues. Thus, Q must be either positive or negative definite. For a minimizer, it must be positive definite, and the solution to the linear system will be the minimizer.

Thus, the quadratic minimization problem with have a unique solution if and only if Q is positive definite, and solving the linear system with a positive definite Q is equivalent to solving the quadratic optimization problem.

6.11

The first iteration of Newton's method gives $x_1 = x_0 - \frac{f\prime(x_0)}{f\prime\prime(x_0)}$. Since $f\prime(x) = 2ax + b$ and $f\prime\prime(x) = 2a$, we see that

$$f\prime(x_1) = 2a(x_0 - [(2ax_0 + b)/(2a)]) + b = 2ax_0 - 2ax_0 - b + b = 0$$

Also, we see that

$$f\prime\prime(x_1) = 2a > 0$$

Therefore, the first derivative is 0 and the second is positive, so this is a minimizer. Since the function is quadratic, we know that this minimizer is unique.

6.15

See secant_method.ipynb for this problem.

7.1

Prop 7.1.5 states that if S is a nonempty subset of V, then $conv(S)$ is convex.

Take $a, b \in conv(S)$. Then $a = \zeta_1 x_1 + ... + \zeta_n x_n$ and $b = \gamma_1 x_1 + ... + \gamma_n x_n$ where $\{x_i\}_{i=1}^n, \{x_i\}_{j=1}^m \in S$ and $\sum_{i=1}^n \zeta_i = \sum_{j=1}^m \gamma_j = 1$ with all $\zeta_i, \gamma_j \in [0, 1]$.

The convex combination can be written as as $\lambda a + (1 + \lambda)b$. Thus,

$$\lambda a + (1 + \lambda)y = \lambda(\zeta_1 x_1 + \cdots \zeta_k x_k) + (1 - \lambda)(\gamma_1 x_1 + \cdots \gamma_k x_k)$$

$$= (\lambda\zeta_1 + (1 - \lambda)\gamma_1)x_1 + \cdots + (\lambda\zeta_k + (1 - \lambda)\gamma_k)x_k$$

The convex combination of a and b are contained in $conv(S)$ because

$$\sum_{i=1}^k (\lambda\zeta_i + (1 - \lambda)\gamma_i) = \lambda \sum_{i=1}^k \zeta_i + (1 - \lambda) \sum_{i=1}^k \gamma_i = \lambda + (1 - \lambda) = 1$$

Thus, Prop 7.1.5 holds.

7.2

(i)

Let $P = \{x \in V : \langle a, x \rangle = b\}$. Let two points $x, y \in P$, such that $a_1 x_1 + \cdots + a_n x_n = b$ and $a_1 y_1 + \cdots + a_n y_n = b$ and $0 \leq \lambda \leq 1$. Then

$$\lambda x + (1 - \lambda)y = \lambda(a_1 x_1 + \cdots + a_n x_n) + (1 - \lambda)a_1 y_1 + \cdots + a_n y_n = \lambda b + (1 - \lambda)b = b$$

Thus, this convex combination also lies on the hyperplane, and the hyperplane is convex.

(ii)

Let $H = \{x \in \mathbb{R}^n : \langle a, x \rangle \leq b\}$. Let two points $x, y \in H$, such that $a_1x_1 + \cdots + a_nx_n = b$ and $a_1y_1 + \cdots + a_ny_n = b$ and $0 \leq \lambda \leq 1$. Then

$$\lambda x + (1-\lambda)y = \lambda(a_1x_1 + \cdots + a_nx_n) + (1-\lambda)a_1y_1 + \cdots + a_ny_n \leq \lambda b + (1-\lambda)b = b$$

Thus, this convex combination also lies in the half space, and a half space is convex.

7.4

(i)

This proof is just algebra. We know $||x-y||^2 = \langle x-y, x-y \rangle = \langle x-p+p-y, x-p+p-y \rangle$

Let $z = x - p$ and $k = p - y$. Then we have,

$$||x-y||^2 = \langle z+k, z+k \rangle = \langle z, z \rangle + 2\langle z, k \rangle + \langle k, k \rangle = \langle x-p, x-p \rangle + \langle p-y, p-y \rangle + 2\langle x-p, p-y \rangle = ||x-p||^2 + ||p-y||^2 + 2\langle x-p, p-y \rangle$$

(ii)

Given (i) and (7.14), we know $\langle x-p, p-y \rangle \geq 0$, therefore:

$$||x-y||^2 - ||x-p||^2 + ||p-y||^2 \geq 0$$

This implies

$$||x-y||^2 + ||p-y||^2 \geq ||x-p||^2$$

When $y \neq p$, this implies

$$||x-y|| > ||x-p||$$

(iii) Given (i) again, we have:

$$||x-x||^2 = ||x-p||^2 + ||p-z||^2 + 2\langle x-p, p-z \rangle = |x-p||^2 + ||p-\lambda y + (1-\lambda)p||^2 + 2\langle x-p, p-(\lambda y + (1-\lambda)p) \rangle = ||x-p||^2 + 2\lambda\langle x-p, p-y$$

(iv)

From (ii) we know that $||x-y||^2 > ||x-p||^2$, and rearranging (7.15) we get:

$$0 \leq 2\langle x-p, p-y \rangle + \lambda^2||y-p||^2$$

7.8

$f : \mathbb{R}^m \to \mathbb{R}$ is convex, $A \in M_{mxn}(\mathbb{R})$, and $b \in \mathbb{R}^m$. Let $x, y \in \mathbb{R}^n$, $x \neq y$, and $\lambda \in [0, 1]$.

Then, we see that $g(\lambda x + (1-\lambda)y) = f(\lambda Ax + (1-\lambda)Ay + b) = f(\lambda(Ax+b) + (1-\lambda)(Ay+b))$.

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

Thus, we can see that $g(x) = f(Ax+b)$ is convex.

7.12

(i)

Take $X, Y \in PD_n(\mathbb{R})$ as positive-definite matrices in $M_n\mathbb{R}$ and $\lambda \in [0,1]$. Then, because X and Y are positive-definite, for every $v \in \mathbb{R}^n$ we have:

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

Thus, since λ is positive, we know the linear combination of these matricies lies within the set, and also lies within the set of positive-definite matrices. Thus, the set is convex.

7.13

If f is bounded by some M, that is, $f(x) < M \forall x$. Assume that f is convex but not constant. Then there would exist $x_1, x_2 \in \mathbb{R}^n$ such that $f(x_1) \neq f(x_2)$. However, if this were true and f was convex and bounded above, f(x) must lie on or above the line between $f(x_1)$ and $f(x_2)$. Because this line must intersect $f(x) = M$, we have a contradiction. Thus, f must be constant.

7.20

Prop 7.4.3 states that if $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $-f$ is also convex, then f is affine.

Since f is convex, we know:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

If $-f$ is also convex, the opposite will hold:

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2)$$

These two equations only hold when the two sides are equal:

$$f(\lambda x_1 + (1-\lambda)x_2) = \lambda f(x_1) + (1-\lambda)f(x_2)$$

Thus, the function is affine and Prop 7.4.3 holds.

7.21

Let $x^* \in \mathbb{R}^n$ be a local minimizer of f. Given this, $f(x^*) \leq f(x)$ for all $x \in \mathcal{B}_r(x^*)$, an open ball around x^* of radius $r > 0$.

Since ϕ is a strictly increasing function, $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{B}_r(x^*)$. Therefore, x^* is a local minimizer of $\phi \circ f$ and the first part holds.

Now let x^* be a local minimizer of $\phi \circ f$. Then $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{B}_r(x^*)$, Again, since ϕ is a strictly increasing function, we know $f(x^*) \leq f(x)$ for all $x \in \mathcal{B}_r(x^*)$. Thus, x^* is a local minimizer of f, and Prop 7.4.11 holds.