You can discuss the problems with each other, but you must write up your answers on your own. Feel free to ask for a hint if you get stuck. [The page/equation numbers are from *Nonlinear Programming*, 2nd edition, 1999.]

The answers to the two * problems are to be turned in jointly with your (randomly chosen) partner.

#1. [Normal cones and simplex constraint]

a) Find the normal cone $N_X(x)$ for the following closed convex sets X:

$$\begin{aligned} X &= \{0\} \subseteq \Re^n, \\ X &= \{x \in \Re^n \mid \|x\| \le 1\}, \\ X &= \{x \in \Re^n \mid a^T x \le b\}, \\ X &= \{x \in \Re^n \mid Ax = d\}, \end{aligned}$$

where $0 \neq a \in \Re^n$, $b \in \Re$, $A \in \Re^{m\times n}$, $d \in \Re^m$ $(n, m \ge 1)$.

b) A simplex constraint can sometimes be easily handled. Find *closed-form* solutions for

$$\min \quad \sum_{j=1}^{n} f_j(x_j) + c_j x_j \quad \text{s.t.} \quad \sum_{j=1}^{n} x_j = 1,$$

with $f_j(x_j) = x_j \ln x_j$, $f_j(x_j) = x_j^2/2$, $f_j(x_j) = \begin{cases} 0 & \text{if } x_j \ge 0 \\ \infty & \text{else} \end{cases}$. You might be able to use the linear constraint to eliminate a variable and make it unconstrained. Simplify your solution if possible.

c) [Exercise 2.1.6] Find closed-form solutions for

$$\max \quad x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \quad \text{s.t.} \quad \sum_{j=1}^{n} x_j = 1, \quad x_j \ge 0, \ j = 1, ..., n,$$

with $a_j > 0$ integer. (Hint: Argue $x > 0$ at any optimal solution and then take $\ln(\cdot)$.) Here a_j is the number of observed occurrences of event j and x_j is the probability of event j occurring, which we estimate by maximum-likelihood.

#2. [Conditional gradient method] Exercise 2.2.1. This is the computational problem for the conditional gradient method. Use starting point $x^0 = (.2, .3, .5)$. [$\bar{x}^k$ has a closed form expression, as does the minimizing stepsize α^k. In particular, if f is a convex quadratic function, then $\phi_k(\alpha) = f(x^k + \alpha d^k)$ is also a convex quadratic function, and its minimum over $[0, 1]$ can be found in closed form.] Don't forget to verify (via analysis) that $(\frac{1}{2}, \frac{1}{2}, 0)$ is a global minimum, as the question asks.

#3.* [Convergence rate of gradient projection method] Suppose $f : \Re^n \mapsto \Re$ satisfies

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \quad \text{and} \quad (\nabla f(x) - \nabla f(y))^T (x - y) \ge \lambda \|x - y\|^2 \quad \forall x, y \in \Re^n, \qquad (\heartsuit)$$

where $0 < \lambda \le L$. Let X be a nonempty closed convex set in $\Re^n$ and let $[x]^+$ denote the (nearest-point) projection of x onto X.

a) Show that, for any $s > 0$ and any $x, y \in \Re^n$, $\|[x - s\nabla f(x)]^+ - [y - s\nabla f(y)]^+\|^2 \le (1 - 2s\lambda + L^2 s^2)\|x - y\|^2$. [Hint: Use the nonexpansive property of $[\cdot]^+$.]

b) Let x^* be a stationary point of f over X (so that $x^* = [x^* - s\nabla f(x^*)]^+$ for any $s > 0$). Use (a) to show that $\{x^k\}$ generated by the gradient projection method with unity stepsize: $x^{k+1} = [x^k - s\nabla f(x^k)]^+$ satisfies $\|x^{k+1} - x^*\|^2 \le (1 - 2s\lambda + L^2 s^2)\|x^k - x^*\|^2$ for $k = 0, 1, ...$ Use this to show that, for $s \in (0, 2\lambda/L^2)$, the rate of convergence is (at least) linear.

c) Suppose $f(x) = \frac{1}{2} x^T Q x + c^T x$ for some symmetric positive definite $Q \in \Re^{n\times n}$ and some $c \in \Re^n$. Show that f satisfies $(\heartsuit)$.

#4. [Gradient-projection for dual] Consider the problem of finding a point $x \in \Re^n$ inside a polyhedral set $\{x | Ax \le b\}$ that is nearest to a point $\bar{x}$ outside the set:

$$\min \quad \frac{1}{2}\|x - \bar{x}\|^2 \quad \text{s.t.} \quad Ax \le b,$$

where $A \in \Re^{m \times n}, b \in \Re^m$.

a) Define the Lagrangian $L(x, \mu) = \frac{1}{2}\|x - \bar{x}\|^2 + \mu^T(Ax - b)$, where $\mu \in \Re^m$. The dual problem is

$$\max_{\mu \ge 0} \min_x L(x, \mu).$$

Show that the dual problem is equivalent to a convex quadratic program of the form

$$\min \quad \frac{1}{2}\mu^T Q \mu + c^T \mu \quad \text{s.t.} \quad \mu \ge 0,$$

for some positive semidefinite $Q \in \Re^{m \times m}$ and $c \in \Re^m$.

b) Using Matlab or your favorite computer language, implement the gradient projection method (2.30), using the Armijo rule along the projection arc (2.32)-(2.33), to solve the dual problem. The input for your code should be $A, b, \bar{x}$. Take $\mu^0 = 0$. Terminate when $\|\mu^k - [\mu^k - s^k g^k]^+\|/s^k \le 10^{-5}$. Your code should also find the primal solution x. Test your code on the set

$$\{(x_1, x_2) \in \Re^2 \mid x_1 + 2x_2 \le 1,\ -3x_1 + 2x_2 \le 4,\ -7x_1 + x_2 \le 2,\ -2x_1 - x_2 \le 2\}$$

and $\bar{x} = (-3, 3)$. What is the (i) number of iterations, (ii) total number of function evaluations?

c) Can the conditional gradient method be applied to solve the dual problem? Why or why not?

#5.* [Network utility optimization] Many protocols for congestion control of data networks (TCP variants) may be interpreted as implicitly solving the optimization problem

$$\min \sum_{j=1}^{n} f_j(x_j) \quad \text{s.t.} \quad Rx \le c,\ x > 0,$$

where x_j is the data rate of source j, $c \in (0, \infty)^m$ is the vector of capacities for the m links in the network, $R \in \{0, 1\}^{m \times n}$ is the routing matrix ($R_{ij} = 1$ if link i is in the path of source j and $R_{ij} = 0$ else), and f_j, the negative of the utility function, is convex. Examples of f_j are $f_j(x_j) = -w_j \ln x_j$ (TCP Vegas, FAST) and $f_j(x_j) = w_j \frac{x_j^{-p}}{p}$ (TCP Reno, HTCP) with $p > 0$, $w_j > 0$. [Chiang et al., "Layering as optimization decomposition: a mathematical theory of network architectures," Proc. IEEE, Vo. 95, 2007.]

a) Define the Lagrangian $L(x, \mu) = \sum_{j=1}^{n} f_j(x_j) + \mu^T(Rx - c)$, where $\mu \in \Re^m$. The dual problem is

$$\max_{\mu \ge 0} \min_x L(x, \mu).$$

Let $x(\mu)$ denote the x that attains the inner minimum (it's a function of μ). For $f_j(x_j) = -w_j \ln x_j$, find $x(\mu)$ and show that the dual problem is equivalent to an optimization problem of the form

$$\min \quad \sum_{j=1}^{n} f_j^*(-R_j^T \mu) + c^T \mu \quad \text{s.t.} \quad \mu \ge 0,$$

with f_j^* a convex function on $(0, \infty)$ (called the *conjugate function* of f_j) and R_j the jth column of R.

b) Let $h(\mu)$ denote the convex objective function of the second dual problem in (a). Show that $\nabla h(\mu) = c - Rx(\mu)$. (This formula is valid for any strictly convex f_j in fact.)

c) TCP Vegas adjusts x and μ in real time by the formula:

$$\begin{aligned} q(t) &= R^T \mu(t), \\ x_j(t) &= \frac{w_j}{q_j(t)}, \quad j = 1, ..., n, \\ y(t) &= Rx(t), \\ \mu_i(t+1) &= \max\left\{0, \mu_i(t) + \frac{y_i(t)}{c_i} - 1\right\}, \quad i = 1, ..., m, \end{aligned}$$

What kind of method is TCP Vegas?