Miscellaneous Nonlinear Programming Exercises

Henry Wolkowicz *

2–08–21

University of Waterloo
Department of Combinatorics & Optimization
Waterloo, Ontario N2L 3G1, Canada

Contents

1 Numerical Analysis Background

Exercise 1.1 *Pretend you have a computer with base 10 and precision 4 that truncates after each arithmetic operation; for example, the sum of* $24.57 + 128.3 = 152.87$ *becomes* 152.8. *What are the results when* $128.3, 24.47, 3.163,$ *and* 0.4825 *are added in ascending order and descending order in this machine? How do these compare with the correct ("infinite-precision") result? What does this show you about adding sequences of numbers on the computer?*

Exercise 1.2 *Assume you have the same computer as in Exercise 1.1, and you perform the computation* $(\frac{1}{3} - 0.3300)/0.3300$. *How many correct digits of the real answer do you get? What does this show you about subtracting almost identical numbers on the computer? Does the accuracy change if you perform the modified calculation* $(\frac{1}{0.3300})(\frac{1}{3} - 0.3300)$?

*Research supported by The Natural Sciences and Engineering Research Council of Canada. E-mail hwolkowicz@uwaterloo.ca

[0] These exercises are for the course: Numerical Methods for Optimization, given at Scuola Matematica Interuniversitaria, August, 2002.

Exercise 1.3 *Now consider the quadratic equation*

$$x^2 - 100x + 1 = 0.$$

Use the formula for the solution of a quadratic equation to find the roots of this equation. (Assume again that you have the same computer as above.) How many correct digits do you get? Why? Can you improve the accuracy by changing the algorithm?

Exercise 1.4 *Consider the following program:*

h=1/2
x=2/3-h
y=3/5-h
e=(x+x+x)-h
f=(y+y+y+y+y)-h
q=f/e

Explain the value of q on a (binary) computer and on a calculator.

2 Basic Lagrange Multipliers

Exercise 2.1 *Use Lagrange multipliers to find the smallest circle in the plane that encloses the points* $(a, 0), (-a, 0), (b, c)$*, where* $a > 0, c > 0$*. What are the values of the Lagrange multipliers for the optimization problem that you used? Was the Lagrangian minimized by the optimum of the optimization problem?*

Exercise 2.2 *In the xy-plane, sketch the ellipse* $F(x, y) = 3x^2 + 2y^2 - 12x - 12y = 0$ *and find its center* (x_0, y_0)*. Determine the values of c such that the center* (x_c, y_c) *of the circle* $c(x^2 + y^2) - 12x - 12y = 0$ *is interior to the ellipse* $F(x, y) = 0$*. Show that these circles are tangent to the ellipse at the origin. Find c such that* (x_c, y_c) *is closest to* (x_0, y_0)*. Determine c such that* $F(x_c, y_c)$ *has a minimum value.*

Exercise 2.3 *Recall that* λ *is an* eigenvalue *of a symmetric matrix A if there is a vector* $x \neq 0$ *such that* $Ax = \lambda x$*. The vector x is an* eigenvector *of A corresponding to* λ*. Show that if x is n eigenvector of A, the corresponding eigenvalue* λ *is given by the formula* $\lambda = R(x)$*, where R is the* Rayleigh quotient

$$R(x) = \frac{x^t A x}{x^t x} \quad (x \neq 0)$$

of A. Show that $R(\alpha x) = R(x)$ *for all numbers* $\alpha \neq 0$*. Hence, conclude that R attains all of its functional values on the unit sphere* $||x|| = 1$*. Show that the minimum value m of R is attained at a point* x_m *having* $||x_m|| = 1$*. Show that* x_m *is an eigenvector of A and that* $m = R(x_m)$ *is the corresponding eigenvalue of A. Hence m is the least eigenvalue of A. Show that the maximum value M of R is the largest eigenvalue of A and that a corresponding eigenvector* x_M *maximizes R. Show that*

$$m||x||^2 \leq x^t A x \leq M||x||^2,$$

for every vector x. Finally, show that if A is positive definite, then we have the additional inequalities

$$\frac{||x||^2}{M} \leq x^t A^{-1} x \leq \frac{||x||^2}{m},$$

where m and M are respectively the smallest and largest eigenvalues of A.

3 Unconstrained Minimization

Exercise 3.1 *Suppose that $f : \Re^n \to \Re$ is a three times differentiable function, $f \in C^3$.*

1. *Describe the method of steepest descent for minimizing f.*
2. *Suppose that f is a strictly convex quadratic function with minimum at x^*.*
   (a) *If the point $x^{(1)}$ can be expressed as*
   $$x^{(1)} = x^* + \mu s, \tag{1}$$
   where s is an eigenvector of $\nabla^2 f$ with eigenvalue λ and $s \in \Re$, show that: $\nabla f(x^{(1)}) = \mu\lambda s$; and if an exact line search in the direction of steepest descent is assumed, then the method of steepest descent terminates in one iteration.
   (b) *If (1) does not hold, show that*
   $$x^{(1)} = x^* + \sum_{i=1}^{m} \mu_i s_i, \tag{2}$$
   where $m > 1$, and for all i, $\mu_i \neq 0$, and the s_i are eigenvectors of the Hessian corresponding to distinct eigenvectors λ_i. And in this case show that the method does not terminate in one iteration.
3. *Describe a method of conjugate gradients for minimizing f.*
4. *Describe Newton's method for minimizing f. Carefully define the type (and order) of convergence. Assume that x^* is a local minimum and the Hessian of f at x^* is positive definite.* **State** *and* **prove** *the local convergence (and rate) of Newton's method to x^*.*

Exercise 3.2 *Which of the following functions are convex, concave, strictly convex or strictly concave (in x or X):*

1. $$e^x$$
2. $$\min\left\{1 - x^2, 0\right\}$$
3. $$\max\{|x_1|, |x_2|\}$$
4. $$||x|| \quad \textit{Euclidean norm in } \Re^n$$
5. $$\frac{1}{x}\int_0^x F(\alpha)d\alpha, \quad \textit{where } F \textit{ is convex}$$
6. $$\sum_{i=1}^{k} a_i \lambda_i(X)$$

   where $X = X^t$ is a real $n \times n$ matrix with ordered eigenvalues $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ and $a_1 \geq \cdots \geq a_k \geq 0$.

Exercise 3.3 *Scaling a minimization problem is very important, e.g. applying steepest descent to a poorly scaled problem results in poor convergence while the best scaling leads to Newton's method. A quadratic minimization problem could be considered perfectly scaled if its Hessian were the identity.*

If we rescale the variable space by $\hat{x} = Tx$*, then in the new variable space, the quadratic model around* $\hat{x}_+$ *has gradient transformed to* $T^{-T}\nabla f(x)$ *and model Hessian* $T^{-T}H_+T^{-1}$*, where* H_+ *is the model Hessian in the original variable space. Thus, the rescaling that leads to an identity Hessian in the model around* $\hat{x}_+$ *satisfies* $T_+^{-T}H_+T_+^{-1} = I$*, or* $H_+ = T_+^T T_+$.

Exercise 3.4 *Show that the BFGS update is invariant under linear transformations of the variable space.*

Exercise 3.5 *The PSB update of* B *is defined*

$$B_+ = B + \frac{us^T + su^T}{s^T s} - \frac{(u^T s)ss^T}{(s^T s)^2},$$

where $u = y - Bs$*. Show that the PSB update is not invariant under linear transformations of the variable space.*

Exercise 3.6 *Let* S *be a nonempty closed convex subset of* $\Re^n$*, and define* $f\colon \Re^n \mapsto \Re$ *by*

$$f(y) = \inf_{x \in S} \|x - y\|.$$

Prove that f *is convex.*

Exercise 3.7 *Consider Rosenbrock's test function*

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

Find the minimum point for this function from the initial estimate $x^{(1)} = (-1\ 1)^T$*. Use the matlab package as well as the following three methods:*

1. *Newton with approximate linesearch,*
2. *quasi-Newton with BFGS update with approximate linesearch,*
3. *steepest descent with exact line search.*

You can experiment with different approximate linesearches, e.g. backtracking with initial steplength of 1. (Please include well-documented programs.)

Exercise 3.8 *Suppose that* A *is a symmetric* $n \times n$ *matrix. Show that the following are equivalent:*

1. A *is positive definite, i.e.* $x^t Ax > 0,\ \forall x \neq 0$*;*
2. *The determinant of the* n *leading principal submatrices is positive;*
3. *The determinant of all principal submatrices is positive;*
4. A *has a* Cholesky factorization $A = GG^t$*,* G *lower triangular.*

Exercise 3.9 *Let* γ *be a monotone nondecreasing function of a single variable (that is,* $\gamma(r) \leq \gamma(r')$ *for* $r' > r$*) which is also convex; and let* f *be a convex function defined on a convex set* $\Omega \subset \Re^n$*. Show that the composite function* $\gamma(f)$ *defined by* $\gamma(f)(x) = \gamma(f(x))$ *is convex on* Ω*.*

Exercise 3.10 *Let* A *be a symmetric matrix. Show that any two eigenvectors of* A*, corresponding to distinct eigenvalues, are* A*-conjugate.*

Exercise 3.11 *Let* $f(x) = \frac{1}{2}x^tAx - b^tx$ *with*

$$A = \begin{bmatrix} 6 & 13 & -17 \\ 13 & 29 & -38 \\ -17 & -38 & 50 \end{bmatrix}, \; b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \; x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

1. *Show that* f *is strictly convex and find the global minimum using the optimality conditions.*
2. *Apply the conjugate gradient algorithm with* x_0 *as the initial point and using first 5 and then 10 significant digits in the arithmetic. In each case, apply an additional fourth step of the CG algorithm. Do you see a significant improvement? Why?*

Exercise 3.12 *Let* $H_n = \left(\frac{1}{i+j-1}\right)$, $i, j = 1, \ldots n$, *denote the* $n \times n$ *Hilbert matrix. This matrix is positive definite but ill-conditioned. Let* $A = I + H_{16}$, *where* I *is the identity matrix. Choose b so that the minimum of* $f(x) = \frac{1}{2}x^tAx - b^tx$ *is the vector of ones. Now apply the CG algorithm to minimize* f. *Comment on the accuracy of the solution after 5,9, and 16 steps.*

Exercise 3.13 *Find the minimum of the following functions using the matlab routine for minimization a) without derivatives, b) with derivatives.*

1. *Rosenbrock's:*
$$100(y - x^2)^2 + (1 - x)^2,$$
start at (-1.2,1.0).
2. *Powell's:*
$$(x + 10y)^2 + 5(z - w)^2 + (y - 2z)^4,$$
start at (3,-1,0,1).

Exercise 3.14 *Prove that a quadratic function of n variables,* $f(x) = \frac{1}{2}x^tAx + b^tx + c, A \in \Re^{n \times n}$ *symmetric,* $b \in \Re^n, c \in \Re$ *has a unique minimizer if, and only if, A is positive definite. What is the minimizer? How would you find it?*

Exercise 3.15 *Use the necessary and sufficient conditions for unconstrained minimization to prove that the solution(s) to the linear least-squares problem*

$$\min ||Ax - b||_2, A \in \Re^{m \times n}, x \in \Re^n, b \in \Re^m, \tag{3}$$

is (are) $x_* \in \Re^n$ *satisfying* $(A^tA)x_* = A^tb$. *[Hint: For the case when A doesn't have full column rank, generalize the previous problem to consider positive semidefinite matrices.]*

Exercise 3.16 *Let* f *be a convex function defined on a convex set* $\Omega \subset \Re^n$. *Show that* f *attains a minimum at* $x \in \Omega$ *if, and only if,* $\nabla f(x)^t d \geq 0$, *for all feasible directions* d *at* x.

Exercise 3.17 *Show that an* $n \times n$ *matrix A is symmetric and positive definite if, and only if, it can be written as* $A = GG^T$ *where G is a lower triangular matrix with positive diagonal elements. This representation is known as the* Cholesky factorization *of A.*

Exercise 3.18 *Apply Newton's method to the function*

$$f(x) = \frac{11}{546}x^6 - \frac{38}{364}x^4 + \frac{1}{2}x^2$$

from $x^{(1)} = 1.01$. *Verify that the Hessian* $G^{(k)}$ *is always positive definite and that* $f^{(k)}$ *is monotonic decreasing. Show nonetheless that the accumulation points of the sequence* $\{x^{(k)}\}$ *are* ± 1, *and that the gradient does not converge to* 0. *Verify that for any fixed positive* ρ *(the constant for sufficient decrease), however small, the Armijo-Goldstein conditions are violated for sufficiently large* k.

Exercise 3.19 *Show that the steepest descent method is invariant under an orthogonal transformation of variables.*

Exercise 3.20 *Consider the trust region subproblem:*

$$(TR) \quad \min q(x) := x^t Q x - 2c^t x \quad \textit{subject to } ||x|| \leq \delta,$$

where $Q = Q^t$, $c \in \Re^n$, and $\delta > 0$. A recent theorem states that a dual problem for (TR) without any duality gap, is the concave maximization problem

$$\max_{\lambda} \quad h(\lambda) := \lambda\delta^2 - c^t(Q - \lambda I)^{-1} c$$

$$\textit{subject to } \; Q - \lambda I \textit{ positive definite}, \; \lambda \leq 0.$$

1. *Write a matlab program to solve (TR), to a given tolerance, by solving the dual program. More precisely:*

   start with an initial estimate for λ that guarantees feasibility, i.e. positive definiteness;

   use the optimality conditions to find an estimate for the optimal solution x;

   check feasibility and optimality with a given tolerance;

   use Newton's method to find the search direction (in $\Re$) to maximize the function h;

   use a backtracking strategy to ensure that positive definiteness is maintained;

   repeat the Newton iteration.

2. *Generate and solve 2 random (TR) problems of size $n = 5, n = 6$ using your matlab program. (Use the function "rand" in matlab to generate the random problems. Include the data that you used.)*

3. *Use your matlab program to solve the (TR) problem with*

$$Q = \begin{bmatrix} 2 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \; c = (1 \; 0 \; 1)^t, \; \delta = \sqrt{3}.$$

   Can you explain the change in the number of iterations? Can you get 10 places accuracy for the solution vector x?

Exercise 3.21 *Consider the following two-variable problem.*

$$\max f(x) = 2x_1x_2 + 2x_2 - 2x_2^2.$$

1. *Perform two steps of steepest descent (by hand). Use the origin as the starting point.*

2. *Use the optimization toolkit in matlab to find the maximum and verify this using the optimality conditions.*

4 Nonlinear Equations

Exercise 4.1 *For each of the functions $f(x) = x, f(x) = x^2 + x, f(x) = e^x - 1$, answer the following questions:*

1. *What is $f'(x)$ at the root 0.*

2. *What is a Lipschitz constant for f' in the interval $[-a, a]$; i.e. what is a bound on $|f'(x) - f'(0))/x|$ in this interval?*

3. *What region of convergence of Newton's method on $f(x)$ is predicted by the convergence theorem?*

4. In what interval containing 0 is Newton's method on $f(x)$ actually convergent to 0?

Exercise 4.2 *Use the answers from the above problem and consider applying Newton's method to finding the root of*

$$F(x) = \begin{bmatrix} x_1 \\ x_2^2 + x_2 \\ e^{x_3} - 1 \end{bmatrix}$$

1. What is the Jacobian $J(x)$ at the root $(0,0,0)^t$?
2. What is a Lipschitz constant on $J(x)$ in an interval of radius A around the root?
3. What is the predicted region of covergence for Newton's method on $F(x)$?

Exercise 4.3 *Attempt to solve*

$$0 = f(x) = (2x + 4ax^3)e^{x^2+ax^4}$$

with

1. $a = 10, x_0 = 1$;
2. $a = 10, x_0 = 2, 3, 10$.

Use Newton's method and a reasonable tolerance. Do not use safeguarding or a linesearch. (You can cancel the exponential term in the Newton iteration to avoid overflow in matlab.) Why is there such a difference in the iteration count? Note that $f(x)$ is the derivative of $e^{x^2+ax^4}$.

Exercise 4.4 *The system of equations*

$$\begin{aligned} 2(x_1 + x_2)^2 + (x_1 - x_2)^2 - 8 &= 0 \\ 5x_1^2 + (x_2 - 3)^2 - 9 &= 0 \end{aligned}$$

has a solution $(1,1)^t$. Carry out one iteration of Newton's method from the starting point $(2,0)^t$. Then solve the system using matlab with the same starting point.

5 Convex Analysis

Exercise 5.1 *Let C be a closed convex set in $\Re^n$. Let $y \in \Re^n \backslash C$. State the basic separation theorem.*

1. Let z be a point on the boundary of C. State the supporting hyperplane theorem. (Do you need some extra assumption on C ?)
2. Let C_1, C_2 in $\Re^n$ be convex sets such that $\text{int}(C_1) \neq \emptyset$ and $\text{int}(C_1) \cap C_2 = \emptyset$. Prove that there exist $c \in \Re^n \backslash \{0\}$, $\beta \in \Re$ such that

$$c^t x \ \leq \ \beta \ \leq c^t y \ \ \forall x \in C_1, \ \forall y \in C_2.$$

Exercise 5.2 *Let C be a closed convex set in $\mathbb{R}^n$. Assume throughout this question that C has interior points.*

1. State the basic separation theorem.
2. Using the basic separation theorem, prove that " if $\bar{x}$ is in the boundary of C then there exists $a \in \mathbb{R}^n$ $a \neq 0$, such that $a^T x \ \leq \ a^T \bar{x}$ for all $x \in C$".
3. Given a convex function $f : C \to \mathbb{R}$, use the previous part of this question to prove that for every interior point $\hat{x}$ of C there exists $a \in \mathbb{R}^n$ such that

$$f(x) \ \geq \ f(\hat{x}) \ + \ a^T(x - \hat{x}), \ \forall x \in C.$$

4. *Let $\hat{x}$ be an interior point of C and f be a convex function on C as in the previous part. Define what is meant by the* subdifferential *of f at $\hat{x}$. Write down necessary and sufficient conditions for optimality of $\hat{x}$ for minimizing f over C (in terms of the subdifferential).*

Exercise 5.3 *Find the dual cones for each of the following cones*

1. $$x \geq 0, \ y \geq 0 \ in \ \Re^2,$$

2. $$x + y \geq 0, \quad 2x + 3y \geq 0, \ in \ \Re^2,$$

3. $$2x + 3y \geq 0, \ in \ \Re^2,$$

4. $$2x + 3y = 0, \ in \ \Re^2,$$

5. *the set of real, $n \times n$, positive semidefinite symmetric matrices.*

Exercise 5.4 *Let K be a closed cone in $\Re^n$. Show that the cone K does not contain a line (called a pointed cone)* if and only if *there is a vector y such that $y^t x < 0$, for every $x \neq 0$ in K.*

Exercise 5.5 *Prove that a supporting hyperplane of a convex compact set K in $\Re^n$ contains at least one extreme point of K.*

Exercise 5.6 *Suppose that C is a convex set in $\Re^n$ and the point $y \notin \bar{C}$, the closure of C. Prove that there exists ϕ such that*

$$\phi^t y < \phi^t x, \quad \forall x \in C.$$

(Hint: First prove that there is a point $\bar{x} \in \bar{C}$ which is closest to y.)

Exercise 5.7 *Suppose that K_1, K_2 are convex sets in $\Re^n$ and the interior of K_1 is not empty, $int(K_1) \neq \emptyset$. In addition, suppose that*

$$K_2 \cap int(K_1) = \emptyset.$$

Prove that there exists a hyperplane separating K_1, K_2.

Exercise 5.8 *Suppose that $S \subset \Re^n$. Prove that $S^{++} = S$ if and only if S is a closed convex cone. ($\cdot^+$ denotes polar)*

6 Constrained Optimization: Optimality Conditions

Exercise 6.1 *Consider the program*

$$\begin{array}{lll} & \max & \frac{1}{3}\sum_{i=1}^n x_i^3 \\ (P) & \text{subject to} & \sum_{i=1}^n x_i = 0 \\ & & \sum_{i=1}^n x_i^2 = n, \end{array}$$

where $n > 2$.

1. *Write down the Lagrangian dual for (P).*

2. *Write down the KKT conditions for (P). Justify that they are satisfied at an optimum. (Show that an optimum exists first and then use an appropriate constraint qualification.)*
3. *Find the largest value of the objective function that occurs at a KKT point and give all the corresponding KKT points.*
4. *State the second order sufficient optimality conditions. Verify that the points satisfying the second order conditions are all local maximizers.*

Exercise 6.2 *Under what conditions on the problem are the KKT conditions*

1. *necessary*
2. *sufficient*
3. *necessary* and *sufficient*

for the solution of an inequality constrained optimization problem?

Form the KKT conditions for the problem

$$\begin{array}{ll} \textit{maximize} & (x+1)^2 + (y+1)^2 \\ \textit{subject to} & x^2 + y^2 \le 2 \\ & y \le 1 \end{array}$$

and hence determine the solution.

Exercise 6.3 *Consider the convex programming problem*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & g_j(x) \le 0, \quad j = 1, \dots m, \end{array}$$

where the functions are all real valued convex functions on $\Re^n$. Let $S \subset \Re^n$ denote the feasible set for the problem. Show that if $\bar{x} \in S$ and none of the functions are identically 0 on the feasible set, then $\bar{x}$ is a regular point for the problem.

Exercise 6.4 *Let P be a symmetric $n \times n$ matrix and Q a positive semi-definite symmetric $n \times n$ matrix. Assume that $x^t P x > 0$, for all $x \neq 0$ satisfying $x^t Q x = 0$. Show that there exists a scalar c such that*

$$P + cQ \succ 0 \quad (\textit{positive definite}).$$

Exercise 6.5 *Consider the following problem:*

$$\begin{array}{ll} \textit{Minimize} & \frac{4}{3}(x_1^2 - x_1 x_2 + x_2^2)^{\frac{3}{4}} - x_3 \\ \textit{subject to} & -x_1, -x_2, -x_3 \le 0 \\ & x_3 \le 2 \end{array}$$

Show that the objective function is convex and that the optimal solution is attained at the unique point $\overline{x} = (0,0,2)^t$. Let the initial point be $x_1 = (0, a, 0)^t$, where $a \le 1/(2\sqrt{2}$. Find feasible directions using the program:

$$\begin{array}{ll} \textit{Minimize} & \nabla f(x_k)^t d \\ \textit{subject to} & \hat{A}_1 d \le 0 \\ & d^t d \le 1 \end{array}$$

where $\hat{A}_1$ is the matrix whose rows are the gradients of the active constraints at x_1. Show that Zoutendijk's method with exact line search yields the sequence

$$x_k = \begin{cases} \left[0, (\frac{1}{2})^{k-1} a, \frac{1}{2}\sum_{j=0}^{k-2} (\frac{a}{2^j})^{\frac{1}{2}}\right] & \textit{if } k \textit{ is odd}, \ k \ge 3 \\ \left[(\frac{1}{2})^{k-1} a, 0, \frac{1}{2}\sum_{j=0}^{k-2} (\frac{a}{2^j})^{\frac{1}{2}}\right] & \textit{if } k \textit{ is even} \end{cases}$$

Does this sequence converge to the optimum? Why? How would you correct the problem?

Exercise 6.6 *Let P be a symmetric $n \times n$ matrix and Q a positive semi-definite symmetric $n \times n$ matrix. Assume that $x^t P x > 0$, for all $x \neq 0$ satisfying $x^t Q x = 0$. Show that there exists a scalar c such that*

$$P + cQ > 0.$$

Exercise 6.7 *Consider the (differentiable) equality constrained program*

$$\begin{array}{rc} Minimize & f(x) \\ subject\ to & h(x) = 0, \end{array}$$

where $h : \Re^n \to \Re^m$. Suppose that the second order sufficient optimality conditions hold at x^*. Show that there exists a scalar $\bar{c}$ such that

$$\nabla^2_{xx} L(x^*, \lambda^*) + \bar{c} \nabla h(x^*) \nabla h(x^*)^t \succ 0 \ (positive\ definite).$$

Use the above to show that x^* is a strict local minimum. (Hint. Use the augmented Lagrangian.)

Exercise 6.8 *Let*

$$\mu_0 = \inf \{ f(x); x \in \Omega,\ g(x) \preceq_S 0,\ h(x) = 0 \},$$

where the definitions coincide with the convex Lagrange multiplier theorem presented in class, and $h(x) = Ax + b$ is an affine map. Assume that μ_0 is finite and that there exists a feasible point $\hat{x}$ such that

$$g(\hat{x}) \prec_S 0,\ h(\hat{x}) = 0,$$

i.e. $g(\hat{x}) \in$ int $(-S)$. (This is called the generalized Slater's constraint qualification. *Show that (Lagrange Multiplier Theorem) there exists Lagrange multipliers λ, μ such that*

$$\mu_0 = \inf \left\{ f(x) + \lambda^t g(x) + \mu^t h(x) : x \in \Omega \right\}.$$

Exercise 6.9 *Let A be an $n \times n$ nonsingular matrix. Prove that*

$$A^{-1} = \lim_{\lambda \to \infty} (\frac{1}{\lambda} I + A^t A)^{-1} A^t.$$

Exercise 6.10 *The* Moore-Penrose generalized inverse *of an $m \times n$ matrix A is the unique $n \times m$ matrix $A^\dagger$ that satisfies the four equations*

$$AA^\dagger A = A,\ A^\dagger A A^\dagger = A^\dagger,\ (AA^\dagger)^t = AA^\dagger,\ (A^\dagger A)^t = A^\dagger A.$$

If A is nonsingular, then $A^\dagger = A^{-1}$. Prove that

$$A^\dagger = \lim_{\lambda \to \infty} (\frac{1}{\lambda} I + A^t A)^{-1} A^t.$$

Exercise 6.11 *Let f be a convex differentiable function of n variables, let A be an (m, n) matrix and b an m -vector. Consider the problem*

$$\min\{ f(x) \mid Ax \leq b, \}. \tag{4}$$

Let x^* be optimal for (4) and assume the gradients of those constraints which are active at x^* are linearly independent. Prove directly that the Karush-Kuhn-Tucker conditions are satisfied at x^*.

7 Duality

Exercise 7.1 *Consider the trust region subproblem (TRS)*

$$\min x^t Ax - 2b^t x \text{ subject to } x^t x - \delta^2 \leq 0,$$

where $A = A^t$. Using the min-max and max-min approach, derive a dual program to (TRS) which is a simple constrained concave maximization program. Then derive the dual to this dual and relate it to the original (TRS) problem.

Exercise 7.2 *Consider the (primal) nonlinear programming problem*

$$\begin{array}{llll} & f^* = & \inf & f(x) \\ (NLP) & & \text{subject to} & g_j(x) \leq 0, \quad j = 1, \dots r, \\ & & & x \in X \subset \Re^n, \end{array}$$

where all functions are assumed to be twice differentiable, there exists at least one feasible solution, and the optimal value is bounded from below, i.e. $-\infty < f^* < \infty$.

A vector $\mu^* = (\mu_1, \dots, \mu_r)$ is said to be a Lagrange multiplier vector *(or simply a* Lagrange multiplier*) for NLP if*

$$\mu_j^* \geq 0, \quad j = 1, \dots, r,$$

and

$$f^* = \inf_{x \in X} L(x, \mu^*).$$

The dual function *is*

$$q(\mu) = \inf_{x \in X} L(x, \mu).$$

The domain of q *is the set where q is finite*

$$D = \{\mu : q(\mu) > -\infty\}.$$

The Lagrangian dual problem *is*

$$q^* = \sup_{\mu \geq 0} q(\mu).$$

Prove the following:

1. *The domain D of the dual function q is convex and Q is concave over D.*
2. *Weak duality holds.*
3. *If there is no duality gap ($q^* = d^*$), then the set of Lagrange multipliers is equal to the set of optimal dual solutions.*
   While, if there is a duality gap, then the set of Lagrange multipliers is empty.

Exercise 7.3 *Under the assumptions and definitions of Exercise 7.2, prove the following:*

1. *(x^*, μ^*) is an optimal solution-Lagrange multiplier pair if and only if*

$$x^* \in X, \quad g(x^*) \leq 0, \qquad \textit{(Primal Feasibility)} \tag{5}$$

$$\mu^* \geq 0, \qquad \textit{(Dual Feasibility)} \tag{6}$$

$$x^* \in \arg\min_{x \in X} L(x, \mu^*), \qquad \textit{(Lagrangian Optimality)} \tag{7}$$

$$\langle \mu^*, g(x^*) \rangle = 0, \qquad \textit{(Complementary Slackness)} \tag{8}$$

2. *(x^*, μ^*) is an optimal solution-Lagrange multiplier pair if and only if $x^* \in X, \mu^* \geq 0$, and (x^*, μ^*) is a saddle point of the Lagrangian, in the sense that*

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \quad \forall x \in X, \mu \geq 0. \tag{9}$$

8 Linear Programming

Exercise 8.1 *Let $a_1, \ldots, m$ be n-vectors, $b_1, \ldots, b_m$ be scalars, $A = [a_1, \ldots, a_m]$, $b = (b_1 \ldots, b_m)'$ and*

$$R = \{x \mid Ax \leq b\}.$$

1. *Define an extreme point for R.*

2. *Assume $R \neq \emptyset$. Prove R possesses an extreme point if and only if rank(A) $= n$.*

Exercise 8.2 *Consider the LP*

$$\min\{c'x \mid a_i'x \leq b_i,\ i = 1, \ldots, m\},$$

where c and $a_1, \ldots a_m$ are n-vectors.

1. *State the Karush-Kuhn-Tucker conditions for this problem.*

2. *Give a constructive proof that these conditions are necessary for optimality provided the gradients of those constraints active at an optimal solution are linearly independent.*

3. *Show geometrically what problems can occur when the linear independence assumption is not satisfied.*

Exercise 8.3 *Let f be a convex differentiable function of n variables and let $\nabla f(x)$ denote its gradient. Let A be an (m, n) matrix and b an m -vector. Consider the problem*

$$\min\{f(x) \mid Ax = b, \quad x \geq 0\}. \tag{10}$$

Show that if $y = x^*$ is optimal for the LP

$$\min\{(\nabla f(x^*))'y \mid Ay = b, \quad y \geq 0\}$$

then x^* solves (10).

9 Penalty and Barrier Methods

Exercise 9.1 *Investigate the Hessian matrices of the log-barrier function. Show that ill-conditioning occurs as the log-barrier parameter converges to 0, if the number of active constraints is between 1 and $n-1$. (Here n is the dimension of the variable space.) What happens when strict complementarity fails at the optimum?*

Exercise 9.2 *Write a program using matlab to find an optimal solution x^* of the problems*

1.

$$\begin{array}{lll} \max & x_1 + x_2 & \\ \textit{subject to} & 1 - x_1^2 - x_2^2 & \geq 0 \\ & x_1 - x_2^2 & \geq 0 \end{array}$$

2.

$$\begin{array}{lll} \max & x_1 - (x_2 + x_3)^2 & \\ \textit{subject to} & x_1 - x_2^2 - 2e^{x_3} & \geq 0 \\ & 10 - x_1 - x_3^4 & \geq 0 \end{array}$$

Use the barrier method. (Compare with the matlab toolkit solution.)

Exercise 9.3 *Write a matlab program to solve the problems*

1.

$$\begin{array}{lrcl} \max & x_1 + x_2 & & \\ \textit{subject to} & x_1^2 + x_2^2 - 1 & = & 0 \end{array}$$

2.

$$\begin{array}{lrcl} \max & x_1 - (x_2 + x_3)^2 & & \\ \textit{subject to} & x_1 + x_2^2 + x_3^2 & = & 0 \\ & x_1 + x_2 + x_3 & \leq & 0 \end{array}$$

Use the penalty function method. (Compare with the matlab toolkit.)

Exercise 9.4 *Consider the optimization problem*

$$\begin{array}{lllll} & \inf & f(x) & & \\ (P) & \textit{subject to} & g_i(x) & \leq & 0, \textit{ for } i \in I, \\ & & h_i(x) & = & 0, \textit{ for } i \in E, \end{array}$$

where f (convex) , g_i (convex), and h_i (affine) are C^1 functions from $\mathbb{R}^n$ to $\mathbb{R}$, I and E are finite index sets. Assume throughout that the problem (P) has optimal solutions.

1. *Write down the Lagrangian of (P) and state the Karush-Kuhn-Tucker Theorem for (P).*
2. *Using a quadratic penalty function, state a family of* **unconstrained** *optimization problems whose optimal solutions (assuming they exist) have as limit points the optimal solutions of (P).*
3. *Reformulate (P) in the form of $(\bar{P})$:*

$$\begin{array}{lllll} & \inf & c^T y & & \\ (\bar{P}) & \textit{subject to} & Ay & = & b, \\ & & y & \in & K, \end{array}$$

   where c and b are vectors, A is a matrix of appropriate dimensions and K is a (possibly nonlinear) convex cone. Describe what A, b, c and K are so that one can get a solution of (P) by solving $(\bar{P})$.
   Hint: *You might want to make the objective function linear first (using a new variable and including f in the constraints). Next, you might want to homogenize your convex set (which involves f, g). Finally, when you express your cone K it should involve some new variables and functions f and g.*

Exercise 9.5 *Solve the problem*

$$\begin{array}{lc} \min & x^2 + xy + y^2 - 2y \\ \text{subject to} & x + y = 2 \end{array}$$

three ways analytically

1. *with the necessary conditions.*
2. *with a quadratic penalty function.*
3. *with an exact penalty function.*

Exercise 9.6 *Consider the problem* $\min f(x)$ s.t. $h(x) = 0$, *where f, h are appropriate functions on $\Re^n$. Show when the condition number of the Hessian of the quadratic penalty function goes to infinity as the penalty parameter goes to infinity. Why should this cause problems in the penalty method?*

Exercise 9.7 *For the problem* $\min f(x)$ s.t. $g(x) \leq 0$, *where* $g(x)$ *is* r *dimensional, define the penalty function*

$$p(x) = f(x) + c \max\{0, g_1(x), g_2(x), \cdots, g_r(x)\}.$$

Let d, μ $(d \neq 0)$ *be a solution to the quadratic program*

$$\begin{array}{ll} \min & \frac{1}{2} d^t B d + \nabla f(x)^t d \\ \text{s.t.} & g(x) + \nabla g(x)^t d \leq 0, \end{array}$$

where B *is positive definite. Show that* d *is a descent direction for* p *for sufficiently large* c.

Exercise 9.8 *Consider the equality constrained nonlinear problem*

$$\begin{array}{llll} & f^* = & \inf & f(x) \\ (NEP) & & \textit{subject to} & h_i(x) = 0, \quad i = 1, \ldots m. \\ & & & x \in X \subset \Re^n, \end{array}$$

where the functions are assumed to be continuous, X *is a closed set, and there exists at least one feasible solution. Define the* augmented Lagrangian function

$$L_c(x, \lambda) = f(x) + \lambda^t h(x) + \frac{c}{2} ||h(x)||^2,$$

where c *is a positive penalty parameter. For* $k = 0, 1, \ldots,$, *let* x^k *be a global minimum of the problem*

$$\begin{array}{lll} f^* = & \min & L_{c^k}(x, \lambda^k) \\ & \textit{subject to} & x \in X, \end{array}$$

where $\{\lambda^k\}$ *is bounded,* $0 < c^k < c^{k+1}$ *for all* k, *and* $c^k \to \infty$. *Then every limit point of the sequence* $\{x^k\}$ *is a global minimum of the original problem NEP.*

10 More Miscellaneous Problems

Exercise 10.1 *Given any* $m \times n$ *matrix* A, *consider the optimization problem*

$$\alpha = \sup\{x^T A y : ||x||^2 = 1, ||y||^2 = 1\} \tag{11}$$

and the matrix

$$\tilde{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}.$$

1. *If* μ *is an eigenvalue of* $\tilde{A}$, *prove* $-\mu$ *is also.*
2. *If* μ *is a nonzero eigenvalue of* $\tilde{A}$, *use a corresponding eigenvector to construct a feasible solution to problem (11) with objective value* μ.
3. *Deduce* $\alpha \geq \lambda_{max}(\tilde{A})$ *(the largest eigenvalue).*
4. *Use the Karush-Kuhn-Tucker theorem to prove any optimal solution of problem (11) corresponds to an eigenvector of* $\tilde{A}$. *(State carefully the:* **theorem, assumptions, constraint qualifications***, that you use, i.e. justify the use of the theorem.)*
5. *Deduce* $\alpha = \lambda_{max}(\tilde{A})$. *(This number is called the* largest singular value of A.*)*

Exercise 10.2 *Consider the convex program*

$$\begin{array}{lll} (CP) & \min & f(x) \\ & \textit{subject to} & g(x) \leq 0, \end{array}$$

where $f\colon \Re^n \to \Re$ *and* $g\colon \Re^n \to \Re^m$ *are differentiable and convex functions on* $\Re^n$.

1. *State Slater's constraint qualification (CQ) for (CP).*

2. *Suppose that Slater's CQ holds and that*

$$x(\alpha) = \arg\min_x \{f(x) - \alpha \sum_{j=1}^{n} \log(-g_j(x)) \mid g(x) < 0\}$$

for $\alpha > 0$. What should α satisfy in order that

$$f(x(\alpha)) - \inf_x \{f(x) \mid g(x) \leq 0\} \leq 10^{-6}$$

Prove your claim.

*(**Hint:** Recall that the primal-dual interior-point approach for convex programming is similar to that for linear programming.)*

Exercise 10.3 *Let f be a convex function from $\Re^n$ to $\Re$. Suppose that for each nonzero y in $\Re^n$ there exists some positive η such that $f(\eta y) > f(0)$.*

1. *Prove that the set*

$$L(1) = \{\, x \in \Re^n \mid f(x) \leq 1 \,\}$$

is compact.

2. *Can you draw any conclusion about the compactness of*

$$L(n) = \{\, x \in \Re^n \mid f(x) \leq n \,\},$$

where n is an arbitrarily large natural number? Explain.

Exercise 10.4 *Let f be a convex differentiable function of n variables, let A be an (m, n) matrix and b an m -vector. Consider the problem*

$$\min\{f(x) \mid Ax \leq b, \}. \tag{12}$$

Let x^* be optimal for (12). Consider the following solution algorithm for (12). Let x_0 be an arbitrary feasible point for (12) and let y_0 be optimal for the LP

$$\min\{\nabla f(x_0)' y \mid Ay \leq b, \}. \tag{13}$$

Let

$$\sigma_0 = \arg\min\{f(x_0 + \sigma(y_0 - x_0) \mid 0 \leq \sigma \leq 1\}.$$

Set $x_1 = x_0 + \sigma_0(y_0 - x_0)$. Replace x_0 with x_1 and repeat.

1. *Sketch the progress of this algorithm.*

2. *Would this algorithm be computationally efficient? Why?*

3. *Verify the bound*

$$f(x^*) \geq f(x_o) + \nabla f(x_0)'(y_0 - x_0).$$

Exercise 10.5 1. *Prove that any local minimum of a convex set is also a global minimum. Do not make any differentiability assumptions,*

2. (a) *State the two term Taylor's series for a twice differentiable function $f(x)$ (x is an n-vector) about a point x_0.*

(b) *By taking gradients of both sides of this Taylor's series, can one conclude that*

$$\nabla f(x) = \nabla f(x_0) + H(\xi)(x - x_0),$$

where ξ lies on the line segment joining x and x_0?

Exercise 10.6 *Let $f : \Re^n \to \Re$ be a differentiable convex function on $\Re^n$ and let $x_1, x_2, \ldots, x_k$ be k points in $\Re^n$ such that for some $u \in \Re^k$:*

$$\sum_{i=1}^{k} u_i \nabla f(x_i) = 0, \quad \sum_{i=1}^{k} u_i = 1, \quad u \geq 0.$$

Derive a lower bound for $\inf_{x \in \Re^n} f(x)$ in terms of $x_1, x_2, \ldots, x_k$ and u.

Exercise 10.7 *Find an upper bound, in terms of the problem data, to the maximum value of the following linear program:*

$$\max p^T x \text{ subject to } Ax \leq b, e^T x \leq 1, x \geq 0.$$

Here $p \in \mathrm{IR^m}, \mathrm{A} \in \mathrm{IR^{m \times n}}, \mathrm{b} \in \ \mathrm{RR^m}$, and e is a vector of ones in $\mathrm{IR^m}$.

Exercise 10.8 *Consider the optimal control problem*

$$\begin{array}{rl} \min & \sum_{i=0}^{N-1} |u_i|^2 \\ \textit{subject to} & x_{i+1} = A_i x_i + B_i u_i, \quad i = 0, \ldots, N-1, \\ & x_0, \quad \textit{given} \\ & x_N \geq c, \end{array}$$

where c is a given vector, and A_i, B_i are matrices of appropriate size. Show that a dual problem is of the form

$$\begin{array}{rl} \min & \mu^T Q \mu + \mu^T d \\ \textit{subject to} & \mu \geq 0, \end{array}$$

where Q is an appropriate $n \times n$ matrix (n is the dimension of x_N) and $d \in \mathrm{IR^n}$ is an appropriate vector.

Exercise 10.9 *For the problem $\min\{f(x)\}$, suppose an iterative method is used according to $x_{j+1} = x_j - B_j g_j$, where $g_j = \nabla f(x_j)$ and B_j is a positive definite matrix chosen to approximate $H_j \equiv H_f(x_j)^{-1}$. Assume $\{x_j\}$ converges to z where $\nabla f(z) = 0$. Let $e_j = \|B_j - H_j\|$. Under what conditions on e_j will the rate of convergence of $\{x_j\}$ be*

1. *superlinear,*
2. *quadratic?*

In each case, prove your result and state any differentiability requirements on f.

Exercise 10.10 *Suppose that for some real number $\gamma > 0$, $x(\gamma) > 0$ solves the interior penalty (barrier) problem*

$$\min \left\{ f(x) - \gamma \sum_{j=1}^{n} \log x_j \;\middle|\; Ax = b \right\}$$

where $f\colon \mathrm{IR^n} \mapsto \mathrm{IR}$, A is an mxn real matrix, b is an mx1 real vector and f is convex and differentiable on $\mathrm{IR^n}$. Give a lower bound to

$$\inf \{f(x) \mid Ax = b, \;\; x \geq 0\}$$

in terms of $f(x(\gamma))$, γ and n. Establish your claim.

Exercise 10.11 *Suppose the problem*

$$\begin{array}{ll} \min & f(x) \\ \textit{subject to} & h(x) = 0, \end{array} \tag{14}$$

(where $f\colon \mathrm{I\!R^n} \mapsto \mathrm{I\!R}$ and $h\colon \mathrm{I\!R^n} \mapsto \mathrm{I\!R^m}$ are continuous functions) has a solution x^*. Let M be an optimistic estimate of $f(x^*)$, that is, $M \leq f(x^*)$. Consider the unconstrained problem

$$\min_x v(M, x) := (f(x) - M)^2 + \|h(x)^2\|. \tag{15}$$

Consider the following algorithm. Given $M_k \leq f(x^*)$, a solution x_k to problem (15) *with $M = M_k$ is found, then M_k is updated by*

$$M_{k+1} = M_k + [v(M_k, x_k)]^{1/2} \tag{16}$$

and the process repeated.

1. *Show that if $M = f(x^*)$, a solution of* (15) *is a solution of* (14).
2. *Show that if x_M is a solution of* (15)*, then $f(x_M) \leq f(x^*)$.*
3. *Show that if $M_k \leq f(x^*)$ then M_{k+1} determined by* (16) *satisfies $M_{k+1} \leq f(x^*)$.*
4. *Show that $M_k \longrightarrow f(x^*)$.*

Exercise 10.12 *Consider the following bounded-variable LP with a single general linear constraint:*

$$\begin{array}{ll} \max_x & \sum_{i=1}^n c_i x_i \\ \textit{subject to} & \sum_{i=1}^n a_i x_i = b \\ & 0 \leq x_i \leq u_i, \quad i = 1, \ldots, n. \end{array}$$

1. *State an LP dual in which there is a dual variable corresponding to the equation and a dual variable for each upper bound constraint.*
2. *Assume that $a_i > 0$ and $u_i > 0, i = 1, \ldots, n$, and that the variables have been indexed so that $c_i/a_i \geq c_{i+1}/a_{i+1}, i = 1, \ldots, n-1$. If $\sum_{i=1}^{n-1} a_i u_i + a_n u_n/2 = b$, state primal and dual optimal solutions and verify that objective values of these two solutions are identical. (Hint: The preceeding expression for b determines a primal optimal solution.)*

Exercise 10.13 *Suppose that we are given Q, a symmetric matrix, and $g \in \Re^n$. Let*

$$q(x) := \frac{1}{2} x^t Q x + g^t x$$

be a quadratic function.

1. *Show that: if q is bounded below, then the minimum of q is attained.*
2. *State two conditions on q that are equivalent to q being bounded below. Then prove this statement, i.e. prove:*

   q is bounded below if and only if (i) <u>. . .</u> and (ii) <u>. . .</u> hold.

Exercise 10.14 *Consider the unconstrained minimization problem*

$$\min\{f(x) : x \in \Re^n\}.$$

Let x_c denote the current approximation to the minimum with gradient g_c and approximate Hessian H_c. Let x_+ be the new approximation to the minimum found using an appropriate quasi-Newton method strategy.

1. *State the* secant equation *for the quasi-Newton update H_+..*

2. *Derive the rank one update to H_c, i.e. find the α_c, z_c in the update formula*

$$H_+ = H_c + \alpha_c z_c z_c^t,$$

so that the secant equation holds.

3. *Show that the rank one update is invariant under invertible linear transformations of the variable space.*

Exercise 10.15 *Consider the constrained optimization problem*

$$(NLP) \quad \inf f(x) \quad \text{subject to:} \quad h(x) = 0, \; g(x) \leq 0,$$

where $f : \Re^n \to \Re$, $h : \Re^n \to \Re^k$, and $g : \Re^n \to \Re^m$.

1. *State the first and second order necessary optimality conditions at an optimimum x^*. (Give appropriate constraint qualifications.)*

   For the remaining parts of this question, assume that f and g are convex functions and that we do not have any equality constraints; call this convex optimization problem (P).

2. *Write down the Lagrangian dual, (D), of (P) (use the notation: λ for the dual variables, $h(\lambda)$ for the dual objective function).*

3. *Let $\bar{x}$ and $\bar{\lambda}$ be feasible solutions of (P) and (D) respectively. Prove that if $f(\bar{x}) = h(\bar{\lambda})$ then $\bar{x}$ and $\bar{\lambda}$ are optimal for (P) and (D) respectively.*

4. *Consider the example: $n = 2$, $m = 1$, $f(x_1, x_2) = e^{-x_2}$, $g_1(x_1, x_2) = \|x\|_2 - x_1$. Show that in this case the optimal objective value of (P) is 1 and the optimal objective value of (D) is 0.*
   State conditions which guarantee that: whenever (P) has a finite optimal value, then (D) has an optimal solution and the optimal values of (P) and (D) coincide. Prove that your conditions are sufficient. (You may use results from other questions of this exam.)

5. *Consider the example:*

$$f(x) = \sum_{j=1}^{n} \left(x_j + e^{-x_j} \right),$$

$$g_i(x) = 1 - x_i, \; i \in \{1, 2, \ldots, n\}.$$

   Are your conditions in part (d) satisfied for this problem ? Write down the Lagrangian dual and find its optimal solution. Using the optimal solution of the Lagrangian dual, find an optimal solution of the primal problem.

Exercise 10.16 *Consider the given nonlinear function $F(t, x) = x_1 e^{-x_2 t}$, where $t \in \Re$ and $x \in \Re^2$. You are given data values $d_i = 2.7, \; 1, \; 0.4, \; 0.1$ at the points $t_i = -1, \; 0, \; 1, \; 2$. You have to fit the given function at these points with the given data. To do this you need to*

$$\min_x f(x) := \frac{1}{2} \sum_{j=1}^{4} \left(F(t_j, x) - d_j \right)^2$$

1. *Derive the Gauss-Newton method for the corresponding overdetermined system of nonlinear equations, i.e.:*

   (a) *Calculate the gradient of f; call it $g(x)$.*

   (b) *Calculate the Hessian of f; call it $H(x)$; and express the Hessian using the 2×4 Jacobian matrix; and call the latter $J(x)$.*

(c) *Write down Newton's method for minimization; but then discard the second order terms in the Hessian.*

2. *Carry out one iteration of the Gauss-Newton method from the initial starting point* $(1, 1)$.

3. *Is a line search required for Gauss-Newton? If yes, discuss.*

Exercise 10.17 *Consider the constrained problem*

$$\min_x f(x) \text{ subject to } x \in S,$$

where f is continuous and the set S is robust, *i.e. S is equal to the closure of the interior of S. In other words, it is possible to get to any boundary point by approaching it from the interior. Let $B(x)$ be a barrier function defined on the interior of S. Suppose*

$$S = \{x : g_i(x) \leq 0, i = 1, 2, \ldots, p\}.$$

1. *Define the three properties of a barrier function.*

2. *Give two examples of barrier functions for S.*

3. *Describe the barrier method for solving (P) and prove:* Any limit point of a sequence $\{x_k\}$ generated by the barrier method is a solution to problem (P).

Exercise 10.18 *For continuously differentiable functions $h_1, h_2, \ldots, h_m : \Re^n \to \Re$, assume that the set*

$$S = \{x \in \Re^n : h_j(x) = 0 \ (j = 1, 2, \ldots, m)\}$$

is nonempty, and that for all points x in S the set

$$\{\nabla h_j(x) : j = 1, 2, \ldots, m\}$$

is linearly independent.

(a) *Prove* $\inf\{\|x\| : x \in S\}$ *is attained.*

(b) *Deduce there is a point x in S satisfying*

$$x \in span\{\nabla h_j(x) : j = 1, 2, \ldots, m\}.$$

State clearly any theorems you use.

Exercise 10.19 *For convex functions $f, g_1, g_2, \ldots, g_m : \Re^n \to (-\infty, +\infty]$, consider the convex program*

$$\begin{cases} \inf & f(x) & & \\ \text{subject to} & g_i(x) & \leq & 0 \ (i = 1, 2, \ldots, m) \\ & x & \in & \Re^n. \end{cases}$$

(a) *Write down the Lagrangian dual problem.*

(b) *State and prove the Weak Duality Inequality. Give an example where the inequality is strict, and state a condition guaranteeing equality.*

(c) *Write down the dual problem for the convex program*

$$\begin{cases} \inf & e^{x_1} + 16e^{x_2} & & \\ \text{subject to} & -x_1 - 2x_2 & \leq & 0 \\ & x & \in & \Re^2, \end{cases}$$

and verify the primal and dual values are equal.

Exercise 10.20 *Suppose that $f : \Re \to \Re$ is twice continuously differentiable. Let $a < b$ and suppose f satisfies:*

(a) $f(a) \leq 0,\ f(b) > 0;$

(b) $f'(u) > 0,\ \forall u \in (a, b);$

(c) $f''(u) > 0,\ \forall u \in (a, b).$

1. *Show that if Newton's method for finding a root (a zero) of f is applied with starting point $x_0 = b$, then the following two statements hold:*

   (a) $\{f(x_k)\}$ *is a nonincreasing sequence.*

   (b) *the sequence $\{x_k\}$ converges to the root x^* of f.*

2. *Show or provide a counterexample to the statement: There exists positive constants K, ϵ such that*

$$||x^k - x^*|| \leq K||x^{k+1} - x^k||^2, \quad \textit{whenever } ||x^{k+1} - x^k|| < \epsilon.$$

   What does this say about the the convergence rate of Newton's method? What would change if $f(a) < 0$. (E.g. define quadratic convergence and determine when it is obtained.)

Exercise 10.21 *Consider the simple quadratic problem*

$$\begin{array}{lll} (QP) & \min & 2x^2 + 2xy + y^2 - 2y \\ \textit{subject to} & & x = 0. \end{array}$$

1. *Describe and apply the quadratic (Courant) penalty function method to solve (QP). (Describe carefully the role and the values needed for the penalty parameter.)*

2. *Describe and apply the absolute-value penalty function method to solve (QP). (Describe carefully the role and the values needed for the penalty parameter.)*

3. *The latter penalty function is called an* exact penalty function. *Describe why this is so and what are the advantages and disadvantages of using an exact penalty function.*

Now consider the general constrained problem

$$\begin{array}{lll} & \min & f(x) \\ (NLP) & \textit{subject to} & h(x) = 0 \\ & & g(x) \leq 0. \end{array}$$

1. *Write down the absolute-value penalty function for (NLP). (Hint: Use the* max *function for the inequality constraints.)*

2. *Now suppose that (NLP) has only equality constraints. Suppose that the point x^* is a regular point and satisfies the second-order sufficiency conditions for a local minimum of (NLP). Let $\{\lambda_i\}$ be the corresponding Lagrange multipliers. Show that: for values of the penalty parameter $c > \max_i\{|\lambda_i|\}$, we have that x^* is also a local minimum of the absolute-value penalty function.*

   Hints: Use the primal function

$$p(u) := \min_x \{f(x) : h_i(x) = u_i, \text{ for } i = 1, 2, \dots, m\} \tag{17}$$

   and the special value of $\nabla p(0)$. Relate the problem of minimizing the function

$$p_c(u) := p(u) + c \sum_i^m |u_i|.$$

with the problem of minimizing the penalty function for (17). Then apply the Mean Value Theorem to $p(u)$ to relate p and p_c. Hence conclude that

$$p_c(u) \geq p(u) + (c - \epsilon - \max|\lambda_i|)\sum_i^m |u_i|,$$

for $\epsilon > 0$.

Exercise 10.22 *Consider the convex program*

$$\begin{array}{lll} (CP) & \min & f(x) \\ \textit{subject to} & & x \in \Omega, \end{array}$$

where f is a real valued convex function on $\Re^n$ and Ω is the nonnegative orthant. Suppose that there exists $x \in \Omega$ such that $\nabla f(x) \in$ int Ω.

1. *Prove that the set of optimal solutions of (CP) is a bounded set.*
2. *Can you generalize the result to sets Ω other than $\Re^n_+$? State and prove the broadest generalization you can.*

Exercise 10.23 *Given a subset C of $\Re^n$, functions $g_1, g_2, \ldots, g_m : C \to \Re$, and a vector b in $\Re^m$, consider the primal problem*

$$\left\{ \begin{array}{lrcl} \textit{minimize} & f(x) & & \\ \textit{subject to} & g_i(x) & \leq & b_i \quad (i = 1, 2, \ldots, m), \\ & x & \in & C. \end{array} \right.$$

(a) *Write down the Lagrangian dual problem.*

(b) *Prove the primal optimal value is no less than the dual optimal value.*

(c) *State conditions under which the optimal values are equal.*

(d) *Consider the problem*

$$\left\{ \begin{array}{lrcl} \textit{minimize} & e^y & & \\ \textit{subject to} & x & \leq & b, \\ & \begin{pmatrix} x & y \\ y & z \end{pmatrix} & \in & S^2_+, \end{array} \right.$$

where S^2_+ denotes the cone of 2×2 real positive semidefinite matrices. Calculate the primal and the dual optimal values for all real $b \geq 0$. Comment on your answers.

Exercise 10.24 *For the problem $\min\{f(x)\}$, suppose an iterative method is used according to $x_{j+1} = x_j - B_j g_j$, where $g_j = \nabla f(x_j)$ and B_j is a positive definite matrix chosen to approximate $H_j \equiv H_f(x_j)^{-1}$. Assume $\{x_j\}$ converges to z where $\nabla f(z) = 0$. Let $e_j = \|B_j - H_j\|$. Under what conditions on e_j will the rate of convergence of $\{x_j\}$ be*

1. *superlinear.*
2. *quadratic?*

In each case, prove your result and state any differentiability requirements on f.

Exercise 10.25 *"(Steiner's Problem): Given a triangle in the plane, consider the problem of finding a point whose sum of distances from the vertices of the triangle is minimal. Show that such a point is either a vertex, or else it is such that each side of the triangle is seen from that point at a 120 degree angle (this is known as the Torricelli point)."*

For fixed $a = (a_1, a_2)^{\mathrm{T}} \in \mathbb{R}^2$, define $d_a : \mathbb{R}^2 \setminus \{a\} \to \mathbb{R}$ by

$$d_a(x) = d_a(x_1, x_2) = \operatorname{dist}(x, a).$$

Then

$$\frac{\partial d_a}{\partial x_1}(x) = \frac{d}{dx_1}\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} = \frac{2(x_1 - a_1)}{2\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2}} = \frac{x_1 - a_1}{d_a(x)}.$$

Since our choice of axes is arbitrary, we see that d_a is C^2 on the open set $\mathbb{R}^2 \setminus \{a\}$, with

$$\nabla d_a(x) = \frac{1}{d_a(x)}\begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix}.$$

Now also fix $b = (b_1, b_2)^{\mathrm{T}}$ and $c = (c_1, c_2)^{\mathrm{T}}$ in $\mathbb{R}^2$; put $X := \mathbb{R}^2 \setminus \{a, b, c\}$, and define $f : X \to \mathbb{R}$ by

$$f(x) := d_a(x) + d_b(x) + d_c(x).$$

So X is open and $f(x)$ is C^2. By Fermat's Theorem, if $x^* \in \arg\min(f)$ then $\nabla f(x^*) = 0$. Moreover, since our choice of coordinates was arbitrary, we need only consider $x^* = 0$. So then We are also free to rotate the axe to obtain $a_2 = 0$, and reflect in the x_2-axis (if needed) to get $a_1 < 0$. Then $\|a\| = -a_1$, so we have

Exercise 10.26 *"(Behaviour of Steepest Descent Near a Saddle Point): Let $f = \frac{1}{2}x^{\mathrm{T}}Qx$, where Q is invertible and has at least one negative eigenvalue. Consider the steepest descent method with constant stepsize and show that unless the starting point x^0 belongs to the subspace spanned by the eigenvectors of Q corresponding to the nonnegative eigenvalues, the generated sequence $\{x^k\}$ diverges."*

Assume $Q = Q^{\mathrm{T}}$, so $\nabla f(x) = Qx$. Being symmetric, Q is diagonal in some orthonormal basis. (Moreover, one can show that steepest descent does not change under an orthogonal transformation of the space.) So we can assume

$$Qx = \sum_{i=1}^{n} \lambda_i x_i;$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Let s be the fixed stepsize, so

$$x^{k+1} = x^k - s\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|} = x^k - s\frac{Qx^k}{\|Qx^k\|} = \sum_{i=1}^{n}\left(1 - s\lambda_i/\left\|Qx^k\right\|\right)x_i^k$$

unless $Qx^k = 0$, in which case the procedure terminates.

Assume x^0 is not in the span of eigenvectors associated with nonnegative eigenvalues of Q. So there is some j with $x_j^0 \neq 0$ and $\lambda_j < 0$. Thus

$$x_j^{k+1} = \Bigg(1 - \underbrace{s\lambda_j/\left\|Qx^k\right\|}_{<0}\Bigg)x_j^k;$$

already showing that the procedure does not terminate. If the sequence $\{x_j^k\}$ is bounded, then $s\lambda_j/\left\|Qx^k\right\| \to 0$, but $s\lambda_j$ is constant, so $\left\|Qx^k\right\| \to \infty$. On the other hand, if $\{x_j^k\}$ is unbounded, since $\left|x_j^k\right|$ is strictly increasing we again obtain $\left\|x^k\right\| \to \infty$. (Remark: we did not need invertibility of Q.) □

Exercise 10.27 *"Consider the positive definite quadratic function $f(x) = \frac{1}{2}x^{\mathrm{T}}Qx$ and the steepest descent method with the stepsize α^k chosen by the Goldstein rule. Show that for all k,*

$$f(x^{k+1}) \leq \left(1 - \frac{16\sigma(1-\sigma)Mm}{(M+m)^2}\right)f(x^k).$$

Explain why when $\sigma = 1/2$ this relation yields the result of Prop. 1.3.1. Hint: *Use the result of Exercise 2.9."*

The Goldstein rule means fixing $\sigma \in (0, 1/2)$ and then choosing each α^k to satisfy

$$\sigma \le \frac{f(x^k + \alpha^k d^k) - f(x^k)}{\alpha^k (\nabla f(x^k))^{\mathrm{T}} d^k} \le 1 - \sigma. \qquad [GR]$$

(Then α^k is used in calculating $x^{k+1} := x^k + \alpha^k d^k$.) With the steepest descent method, $d^k = -\nabla f(x^k)$. Assume $Q = Q^{\mathrm{T}}$, so $\nabla f(x) = Qx$ and $f(x+y) = f(x) + x^{\mathrm{T}} Q y + f(y)$. Thus

$$f(x^{k+1}) = f(x^k) + \alpha^k (d^k)^{\mathrm{T}} Q x + f(\alpha^k d^k) = f(x^k) - \alpha^k \left\| d^k \right\|^2 + (\alpha^k)^2 f(d^k). \qquad [1]$$

This allows us to rewrite the middle term of $[GR]$ as follows: ???? That is, $\sigma - 1 \le -\alpha^k f(d^k) / \left\| d^k \right\|^2 \le -\sigma$; so we can rewrite (GR) as

$$\frac{\sigma \left\| d^k \right\|^2}{f(d^k)} \le \alpha^k \le \frac{(1-\sigma) \left\| d^k \right\|^2}{f(d^k)}. \qquad [2]$$

Since $\sigma \le 1/2$, we have

$$\sigma(1-\sigma) \ge \sigma - \sigma^2 + 2\sigma - 1 = \sigma - (1-\sigma)^2. \qquad [3]$$

Since $Q = Q^{\mathrm{T}} \succ 0$, we may apply the Kantorovich inequality (to d^k) and obtain

$$\frac{\left\| d^k \right\|^4}{f(d^k) f(x^k)} = \frac{[(d^k)^{\mathrm{T}} d^k]^2}{\frac{1}{2}[(d^k)^{\mathrm{T}} Q d^k] \frac{1}{2}[(d^k)^{\mathrm{T}} Q^{-1} d^k]} \ge \frac{16Mm}{(M+m)^2}. \qquad [4]$$

Putting the numbered results together, we have as we were to prove. Applying this to the limiting case of $\sigma = 1/2$ gives

$$\frac{f(x^{k+1})}{f(x^k)} \le 1 - \frac{4Mm}{(M+m)^2} = \left(\frac{M-m}{M+m} \right)^2; \qquad [5]$$

which is the same result that Proposition 1.3.1 gives for the line minimization rule; and we are asked why.

Let us assume the context of steepest descent for a positive definite quadratic cost function. Exercise 1.2.9 showed that α^k satisfies the Goldstein rule if and only if $2\sigma\bar{\alpha}^k \le \alpha^k \le 2(1-\sigma)\bar{\alpha}^k$, where $\bar{\alpha}^k$ is the stepsize obtained by the line minimization rule. Taking $\sigma = 1/2$ makes both sides equalities, whence the only α^k satisfying the Goldstein rule is $\alpha^k = \bar{\alpha}^k$. That is, Exercise 1.2.9 implies that the Goldstein rule with $\sigma = 1/2$ simply is the line minimization rule, whence the hypotheses of Proposition 1.3.1 are satisfied. At least in this context, the Goldstein rule is a direct generalization of line minimization, so the answer to 'why?' is 'because Exercise 1.3.5 is a successful generalization of Proposition 1.3.1'. □

Exercise 10.28 *"Consider the simplex method applied to a strictly convex function f. Show that at each iteration, either $f(x_{max})$ decreases strictly, or else the number of vertices x^i of the simplex such that $f(x^i) = f(x_{max})$ decreases by at least one."*

For a simplex iteration in $\mathbb{R}^n$, we have $n+1$ points $S = \{x^0, x^1, \ldots, x^n\} \subset \mathbb{R}^n$, and have chosen $x_{min} \in \arg\min_S f$ and $x_{max} \in \arg\max_S f$. We also define the symbol $\hat{x}$ by

$$\hat{x} := \frac{1}{n}\left(-x_{\max} + \sum_{i=0}^{n} x^i \right)$$

and may as well observe now that Jensen's inequality gives

$$f(\hat{x}) \le \sum_{\substack{i=0 \\ x_i \ne x_{max}}} \frac{f(x^i)}{n} \le f(x_{max}). \qquad [1]$$

The iteration calculates $x_{ref} := 2\hat{x} - x_{max}$ and then considers three cases:
(1) If $f(x_{min}) > f(x_{ref})$, calculate $x_{exp} := 2x_{ref} - \hat{x}$ and put

$$x_{new} := \begin{cases} x_{exp} & \text{if } f(x_{exp}) < f(x_{ref}); \\ x_{ref} & \text{otherwise.} \end{cases}$$

(2) If $\max\{f(x^i):\ x^i \neq x_{max}\} > f(x_{ref}) \geq f(x_{min})$*, put* $x_{new} := x_{ref}$.
(3) If $f(x_{ref}) \geq \max\{f(x^i):\ x^i \neq x_{max}\}$*, put*

$$x_{new} := \begin{cases} \frac{1}{2}(x_{max} + \hat{x}) & \text{if } f(x_{max}) \leq f(x_{ref}), \\ \frac{1}{2}(x_{ref} + \hat{x}) & \text{otherwise.} \end{cases}$$

Then it replaces that x^i *labelled* x_{max} *by* x_{new}*. What we must show is that* $f(x_{new}) < f(x_{max})$*. In case (1), we see that* $f(x_{new}) \leq f(x_{ref}) < f(x_{max})$*. In case (2), we see that* $f(x_{new}) = f(x_{ref}) < \max\{f(x^i):\ x^i \neq x_{max}\} \leq f(x_{max})$*. For case (3), the easier subcase is the second, which only occurs when* $f(x_{ref}) < f(x_{max})$*, and then we see that* $f(x_{new}) = f(\frac{1}{2}x_{ref} + \frac{1}{2}\hat{x}) \leq \frac{1}{2}f(x_{ref}) + \frac{1}{2}f(\hat{x}) < f(x_{max})$*. The first subcase requires a little more care. Suppose failure occurs in the first subcase of (3), so*

$$f(x_{max}) \leq f(x_{new}) = f(\frac{1}{2}x_{max} + \frac{1}{2}\hat{x}) \leq \frac{1}{2}f(x_{max}) + \frac{1}{2}f(\hat{x}) \leq f(x_{max}). \qquad [2]$$

Then $f(\hat{x}) = f(x_{max})$*, implying equality in* [1]*, which can only occur for strictly convex* f *when all points of* $S \setminus \{x_{max}\}$ *coincide. Moreover,* [2] *also implies that* $f(x_{max}) = f(\frac{1}{2}x_{max} + \frac{1}{2}\hat{x}) = f(\hat{x})$*, which can only occur for strictly convex* f *when* $x_{max} = \hat{x}$*. So failure in the first subcase implies that* $x^0 = x^1 = \cdots = x^n$*. Avoiding this pathological event seems beyond the scope of this exercise (since we are not given initial conditions on* S*). Indeed, it could easily be remedied by randomly replacing at least one of the points.* □

Exercise 10.29 *Suppose that* $f : \Re^n \to \Re$.

1. *Define* epigraph of f.
2. *Define* convex function *using the definition of the epigraph.*
3. *Let* $f_i, i \in I$ *be a collection of convex functions defined on a convex set* Ω*. Show that the function* g *defined by* $g(x) = \sup_{i \in I} f_i(x)$ *is convex on the region where it is finite.*

Exercise 10.30 *Consider the problem*

$$\min 5x^2 + 5y^2 - xy - 11x + 11y + 11.$$

1. *Find a point satisfying the first-order necessary conditions for a solution.*
2. *Show that this point is a global minimum.*
3. *What would be the rate of convergence of steepest descent for this problem?*
4. *Starting at* $x = y = 0$*, how many steepest descent iterations would it take (at most) to reduce the function value to* 10^{-11}?

1. *Let* Q *be a symmetric matrix. Show that any two eigenvectors of* Q*, corresponding to distinct eigenvalues, are* Q*-conjugate.*
2. *Show that a conjugate direction method terminates for the minimization of a quadratic function in at most* n *exact line searches.*

Exercise 10.31 *Consider the problem*

$$\begin{array}{ll} \min(\max) & x_1 + x_2^2 + x_2x_3 + 2x_3^2 \\ \textit{subject to} & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) = 1. \end{array}$$

1. *State the first-order necessary conditions for an optimum.*

2. *Verify that $x_1 = 1, x_2 = x_3 = 0, \lambda = -1$ satisfies the first-order conditions. State and check the second-order conditions at this point. Conclude whether this point is a minimum, maximum, or saddle point.*

Exercise 10.32 *Show that the feasible directions program*

$$\begin{array}{ll} \min & g^t d \\ \textit{subject to} & Ad \leq 0 \\ & \max_i |d_i| = 1 \end{array}$$

can be converted to a linear program.

Exercise 10.33 *Suppose that the infinity norm constraint on d above is changed to the Euclidean norm (trust region) constraint*

$$\sum_{i=1}^{n} d_i^2 = 1.$$

Find the Karush-Kuhn-tucker necessary conditions for this problem and show how they can be solved by a modification of the simplex method.

Exercise 10.34 *(Combined penalty and barrier method.) Consider a problem of the form*

$$\begin{array}{ll} \min & f(x) \\ \textit{subject to} & x \in S \cap T. \end{array}$$

Suppose that P is a penalty function for S: (i) P is continuous, (ii) $P(x) \geq 0, \forall x$, and (iii) $P(x) = 0$ if and only if $x \in S$. Suppose that B is a barrier function for T: (i) B is continuous, (ii) $B(x) \geq 0, \forall x$, and (iii) $B(x) \to \infty$ as x approaches the boundary of T.

Define

$$d(c, x) = f(x) + cP(x) + \frac{1}{c}B(x).$$

Let $\{c_k\}$ be a sequence $c_k \to \infty$, and for each k let x_k be a solution to

$$\min d(c_k, x),$$

subject to $x \in$ interior of T. Assume all functions are continuous, T is compact and convex, the original problem has a solution x^*, and that $S \cap$ (interior of T) is not empty. show that

1. $$\lim_{k \to \infty} d(c_k, x_k) = f(x^*).$$

2. $$\lim_{k \to \infty} c_k P(x_k) = 0.$$

3. $$\lim_{k \to \infty} \frac{1}{c_k} B(x_k) = 0.$$

Exercise 10.35 *Solve the quadratic program*

$$\begin{array}{ll} \min & x^2 - xy + y^2 - 3x \\ \textit{subject to} & x \geq 0 \\ & y \geq 0 \\ & x + y \leq 4 \end{array}$$

by use of the active set method starting at $x = y = 0$. Please show your work.

Exercise 10.36 *Consider the constrained optimization problem*

$$(NLP) \quad \inf f(x) \quad \text{subject to:} \quad h(x) = 0, \; g(x) \leq 0,$$

where $f : \Re^n \to \Re$, $h : \Re^n \to \Re^k$, *and* $g : \Re^n \to \Re^m$.

1. *State the first and second order necessary optimality conditions at an optimimum* x^*. *(Give appropriate constraint qualifications.)*

   For the remaining parts of this question, assume that f and g are convex functions and that we do not have any equality constraints; call this convex optimization problem (P).

2. *Write down the Lagrangian dual,* (D), *of* (P) *(use the notation:* λ *for the dual variables,* $h(\lambda)$ *for the dual objective function).*
3. *Let* $\bar{x}$ *and* $\bar{\lambda}$ *be feasible solutions of* (P) *and* (D) *respectively. Prove that if* $f(\bar{x}) = h(\bar{\lambda})$ *then* $\bar{x}$ *and* $\bar{\lambda}$ *are optimal for* (P) *and* (D) *respectively.*
4. *Consider the example:* $n = 2$, $m = 1$, $f(x_1, x_2) = e^{-x_2}$, $g_1(x_1, x_2) = \|x\|_2 - x_1$. *Show that in this case the optimal objective value of* (P) *is 1 and the optimal objective value of* (D) *is 0.*
5. *For a general convex program (P), state conditions which guarantee that: whenever* (P) *has a finite optimal value, then* (D) *has an optimal solution and the optimal values of (P) and (D) coincide.*
6. *Consider the example:*

$$f(x) = \sum_{j=1}^{n} \left(x_j + e^{-x_j} \right),$$

$$g_i(x) = 1 - x_i, \; i \in \{1, 2, \ldots, n\}.$$

   Are your conditions in part (d) satisfied for this problem ? Write down the Lagrangian dual and find its optimal solution. Using the optimal solution of the Lagrangian dual, find an optimal solution of the primal problem.

Exercise 10.37 *Let* X *be a convex set in the* n*-dimensional real space* $\Re^n$ *and let* f *be a differentiable function on* $\Re^n$.

1. *What can you say about the quantity*

$$\nabla f(\bar{x})^T (x - \bar{x})$$

   on X *if* $\bar{x}$ *is a local solution of*

$$\min_{x \in X} f(x).$$

2. *Let* $\bar{x} \in \bar{X}$, *where* $\bar{X}$ *is the solution set of* $\min_x f(x)$ *that is*

$$\bar{X} := \arg\min_{x \in X} f(x).$$

   If f *is convex on* $\Re^n$, *what is the relation between* $\bar{X}$ *and the set*

$$\bar{Y} := \arg\min_{x \in X} \nabla f(\bar{x})^T (x - \bar{x}).$$

Exercise 10.38 *Consider the given nonlinear function* $F(t, x) = x_1 e^{-x_2 t}$, *where* $t \in \Re$ *and* $x \in \Re^2$. *You are given data values* $d_i = 2.7, \; 1, \; 0.4, \; 0.1$ *at the points* $t_i = -1, \; 0, \; 1, \; 2$. *You have to fit the given function at these points with the given data. To do this you need to*

$$\min_x f(x) := \frac{1}{2} \sum_{j=1}^{4} \left(F(t_j, x) - d_j \right)^2$$

1. *Derive the Gauss-Newton method for the corresponding overdetermined system of nonlinear equations, i.e.:*
   (a) *Calculate the gradient of f; call it $g(x)$.*
   (b) *Calculate the Hessian of f; call it $H(x)$; and express the Hessian using the 2×4 Jacobian matrix; and call the latter $J(x)$.*
   (c) *Write down Newton's method for minimization; but then discard the second order terms in the Hessian.*
2. *Carry out one iteration of the Gauss-Newton method from the initial starting point $(1, 1)$.*
3. *Is a line search required for Gauss-Newton? If yes, discuss.*

Exercise 10.39 *Let $f(x) = \frac{1}{2}x^t Qx$, where Q is symmetric, invertible, and has at least one negative eigenvalue. Consider the* steepest descent method *with constant stepsize and show that unless the starting point x^0 belongs to the subspace spanned by the eigenvectors of Q corresponding to the nonnegative eigenvalues, the generated sequence $\{x^k\}$ diverges.*

Exercise 10.40 *Derive the dual (Lagrangian) of the projection problem*

$$\begin{array}{ll} \min & \|z - x\|^2 \\ \textit{subject to} & Ax = 0, \end{array}$$

where the $m \times n$ matrix A and the vector $z \in \Re^n$ are given. Show that the dual problem is also a problem of projection on a subspace.

Exercise 10.41 *Let $f : \Re^n \to \Re$ be a differentiable convex function on $\Re^n$ and let $x_1, x_2, \ldots, x_k$ be k points in $\Re^n$ such that for some $u \in \Re^k$:*

$$\sum_{i=1}^{k} u_i \nabla f(x_i) = 0, \quad \sum_{i=1}^{k} u_i = 1, \quad u \geq 0.$$

Derive a lower bound for $\inf_{x \in \Re^n} f(x)$ in terms of $x_1, x_2, \ldots, x_k$ and u.

Exercise 10.42 CONVEXITY
Suppose that $f : \Re^n \to \Re$.

1. *Define* epigraph of f.
2. *Define* convex function *using the definition of the epigraph.*
3. *Define* supporting hyperplane *to the epigraph of a convex function; and state when such a hyperplane exists.*
4. *Let $f_i, i \in I$ be a collection of convex functions defined on a convex set Ω. Show that the function g defined by $g(x) = \sup_{i \in I} f_i(x)$ is convex on the region where it is finite.*

Exercise 10.43 *MARATOS EFFECT*
This problem illustrates a fundamental difficulty in attaining superlinear convergence using the nondifferentiable exact penalty function for monitoring descent. Consider the problem

$$\begin{array}{ll} \min & f(x) = x_1 \\ \textit{subject to} & h(x) = x_1^2 + x_2^2 - 1 = 0, \end{array}$$

with optimal solution $x^* = (-1, 0)$ and Lagrange multiplier $\lambda^* = \frac{1}{2}$. For any x, let (d, λ) be an optimal solution-Lagrange multiplier pair of the problem

$$\begin{array}{cc} \min & \nabla f(x) + \frac{1}{2} d^t \nabla^2 L(x^*, \lambda^*) d \\ \textit{subject to} & h(x) + \nabla h(x)^t d = 0. \end{array}$$

(Note that d is the Newton direction.) Show that for all c,

$$f(x+d) + c|h(x+d)| - f(x) - c|h(x)| = \lambda h(x) - c|h(x)| + (c - \lambda^*)||d||^2.$$

Conclude that for $c > \lambda^*$, there are points x arbitrarily close to x^* for which the exact penalty function $f(x) + c|h(x)|$ is not reduced by a pure Newton step.

Exercise 10.44 KKT CONDITIONS
Under what conditions on the problem are the KKT conditions:

1. *necessary;*
2. *sufficient;*
3. *necessary and sufficient?*

Form the KKT conditions for the problem

$$\begin{array}{cc} \max & (x+1)^2 + (y+1)^2 \\ \textit{subject to} & x^2 + y^2 \le 2 \\ & y \le 1 \end{array}$$

and hence determine the solution.

Exercise 10.45 NEWTON'S METHOD
Consider the function

$$f(x) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4.$$

Find the largest open ball about $x^* = 0$ in which the Hessian $G(x)$ is positive definite. For what values of the initial estimate x^1 in this ball does Newton's method converge, assuming that $x_1^1 = x_2^1$, i.e. the first coordinates of the initial estimate are equal. At what rate is the convergence? (Define precisely the rate you are using.)

Exercise 10.46 LAGRANGIAN DUALITY

1. *Consider the convex program*

$$\begin{array}{cc} \min & f(x) \\ \textit{subject to} & g(x) \le 0, \\ & x \in X, \end{array}$$

   where $f : \Re^n \to \Re$ is a convex function, X is a convex set, $g : \Re^m \to \Re^m$, $g(x) = (g_i(x))$, and g_i is a convex function for all i. Provide a dual problem. Then state and prove a Lagrangian duality theorem, i.e. provide necessary and sufficient conditions for optimality for the primal and dual problems. (Provide the appropriate constraint qualification when needed.)

2. *Derive the dual (Lagrangian) of the projection problem*

$$\begin{array}{cc} \min & ||z - x||^2 \\ \textit{subject to} & Ax = 0, \end{array}$$

   where the $m \times n$ matrix A and the vector $z \in \Re^n$ are given. Show that the dual problem is also a problem of projection on a subspace.

Exercise 10.47 PSHNENICHNYI CONDITION
Consider the problem

$$\min_{x \in X} := \min_x \{f(x) : Ax \geq b, x \geq 0\},$$

where $f : \Re^n \to \Re$ *is a differentiable* concave *function on* $\Re^n$ *which is bounded below on the nonempty set* X. *Consider the successive linearization algorithm*

$$x^{i+1} \in \textit{arg vertex} \min_{x \in X} \nabla f(x^i)(x - x^i); \quad \textit{stop if } \nabla f(x^i)(x^{i+1} - x^i) = 0.$$

Show that this algorithm is well defined and that $\{x^i\}$ *terminates at some* x^i *which satisfies the minimum principle necessary optimality condition, i.e. the Pshnenichnyi condition.*

Exercise 10.48 STEEPEST DESCENT
Let $f(x) = \frac{1}{2}x^t Qx$, *where* Q *is symmetric, invertible, and has at least one negative eigenvalue. Consider the steepest descent method with constant stepsize and show that unless the starting point* x^0 *belongs to the subspace spanned by the eigenvectors of* Q *corresponding to the nonnegative eigenvalues, the generated sequence* $\{x^k\}$ *diverges.*

Exercise 10.49 KKT CONDITIONS

1. *Under what conditions on the problem are the KKT conditions:*
   (a) *necessary;*
   (b) *sufficient;*
   (c) *necessary and sufficient?*
2. *Form the KKT conditions for the problem*

$$\begin{array}{cc} \max & (x+1)^2 + (y+1)^2 \\ \textit{subject to} & x^2 + y^2 \leq 2 \\ & y \leq 1 \end{array}$$

   and hence determine the solution.

Exercise 10.50 CONVEXITY
Let $f : \Re^n \to \Re$ *be a differentiable convex function on* $\Re^n$ *and let* $x_1, \ldots, x_k$ *be* k *points in* $\Re^n$ *such that for some* $\bar{u} \in \Re^k$:

$$\sum_{i=1}^{k} \bar{u}_i \nabla f(x^i) = 0, \;\; \sum_{i=1}^{k} \bar{u}_i = 1, \;\; \bar{u} \geq 0.$$

Give a lower bound to $\inf_{x \in \Re^n} f(x)$ *in terms of* $x_1, \ldots, x_k$ *and* $\bar{u}$.

Exercise 10.51 LOG-BARRIER
Suppose that for some real number $\mu > 0$, *we have* $x(\mu) > 0$ *solves the log-barrier problem*

$$\min \left\{ f(x) - \mu \sum_{j=1}^{n} log x_j : Ax = b, \right\}$$

where $f : \Re^n \to \Re, A$ *is an* $m \times n$ *real matrix,* b *is an* $m \times 1$ *real vector and* f *is convex and differentiable. Give a lower bound to*

$$\inf \{f(x) : Ax = b, x \geq 0\}$$

in terms of $f(x(\mu), \mu, n$. *Establish your claim.*