Nonlinear optimization
Lecture notes for the course MAT-INF2360
Solutions manual.

Øyvind Ryan

March 5, 2015

Contents

Chapter 1

1. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ with 10 global minima.

2. Consider the function $f(x) = x\sin(1/x)$ defined for $x > 0$. Find its local minima. What about global minimum?

3. Let $f : X \to \mathbb{R}_+$ be a function (with nonnegative function values). Explain why it is equivalent to minimize f over $x \in X$ or minimize $f^2(x)$ over X.
Solution: You can argue in many ways here: For instance the derivative of $f^2(x)$ is $2f(x)f'(x)$, so that extremal points of f are also extremal points of f^2.

4. In Example 1.2.3 we mentioned that optimizing the function $p_x(y)$ is equivalent to optimizing the function $\ln p_x(y)$. Explain why maximizing/minimizing g is the same as maximizing/minimizing $\ln g$ for any positive function g.

5. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(\boldsymbol{x}) = (x_1 - 3)^2 + (x_2 - 2)^2$. How would you explain to *anyone* that $\boldsymbol{x}^* = \mathbf{(3, 2)}$ is a minimum point?

6. The *level sets* of a function $f : \mathbb{R}^2 \to \mathbb{R}$ are sets of the form $L_\alpha = \boldsymbol{x} \in \mathbb{R}^2 : f(\boldsymbol{x}) = \alpha\}$. Let $f(\boldsymbol{x}) = \frac{1}{4}(x_1 - 1)^2 + (x_2 - 3)^2$. Draw the level sets in the plane for $\alpha = 10, 5, 1, 0.1$.

7. The *sublevel set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set $S_\alpha(f) = \{\boldsymbol{x} \in \mathbb{R}^2 : f(\boldsymbol{x}) \leq \alpha\}$, where $\alpha \in \mathbb{R}$. Assume that $\inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n\} = \eta$ exists.

a. What happens to the sublevel sets S_α as α decreases? Give an example.

b. Show that if f is continuous and there is an $\boldsymbol{x}'$ such that with $\alpha = f(\boldsymbol{x}')$ the sublevel set $S_\alpha(f)$ is bounded, then f attains its minimum.

8. Consider the portfolio optimization problem in Subsection 1.2.1.

a. Assume that $c_{ij} = 0$ for each $i \neq j$. Find, analytically, an optimal solution. Describe the set of all optimal solutions.
Solution: If $c_{i,j} = 0$ the function to be minimized is

$$\alpha \sum_{i \leq n} c_{ii} x_i^2 - \sum_{j=1}^{n} \mu_j x_j.$$

The gradient of this function is $2\alpha C\boldsymbol{x} - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the vector with μ in all entries. Lagrange multipliers thus gives that $2\alpha C\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is the vector with λ in all entries. This gives that $x_i = \frac{\mu+\lambda}{2\alpha c_i}$. If $\sum x_i = 1$ we must have that $\frac{\mu+\lambda}{2\alpha}\sum \frac{1}{c_i} = 1$, so that $\lambda = -\mu + \frac{2\alpha}{\sum 1/c_i}$.

b. Consider the special case where $n = 2$. Solve the problem (hint: eliminate one variable) and discuss how minimum point depends on α.
Solution: When $n = 2$, we have that $x_2 = 1 - x_1$, so that

$$\begin{aligned}
f(x_1, x_2) &= \alpha c_{11}x_1^2 + \alpha c_{22}x_2^2 + \alpha(c_{12} + c_{21})x_1x_2 - \mu x_1 - \mu x_2 \\
&= \alpha c_{11}x_1^2 + \alpha c_{22}(1 - x_1)^2 + \alpha(c_{12} + c_{21})x_1(1 - x_1) - \mu x_1 - \mu(1 - x_1) \\
&= \alpha(c_{11} + c_{22} - c_{12} - c_{21})x_1^2 + \alpha(-2c_{22} + c_{12} + c_{21})x_1 + \alpha c_{22} - \mu
\end{aligned}$$

The derivative of this is $2\alpha(c_{11} + c_{22} - c_{12} - c_{21})x_1 + \alpha(-2c_{22} + c_{12} + c_{21})$, which is 0 when $x_1 = -\frac{-2c_{22}+c_{12}+c_{21}}{2(c_{11}+c_{22}-c_{12}-c_{21})}$. This is not dependent on α.

9. Later in these notes we will need the expression for the gradient of functions which are expressed in terms of matrices.

a. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by $f(\boldsymbol{x}) = \boldsymbol{q}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{q}$, where $\boldsymbol{q}$ is a vector. Show that $\nabla f(\boldsymbol{x}) = \boldsymbol{q}$, and that $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{0}$.
Solution: We have that $f(\boldsymbol{x}) = \sum_i q_i x_i$, so that $\frac{\partial f}{\partial x_i} = q_i$, so that $\nabla f(\boldsymbol{x}) = \boldsymbol{q}$. Clearly $\frac{\partial^2 f}{\partial x_i \partial x_j} = 0$, so that $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{0}$.

b. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function $f(\boldsymbol{x}) = (1/2)\boldsymbol{x}^T A\boldsymbol{x}$, where A is symmetric. Show that $\nabla f(\boldsymbol{x}) = A\boldsymbol{x}$, and that $\nabla^2 f(\boldsymbol{x}) = A$.
Solution: We have that

$$f(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j} x_i A_{ij} x_j = \frac{1}{2}\sum_i A_{ii}x_i^2 + \frac{1}{2}\sum_{i,j,i\neq j} x_i A_{ij} x_j,$$

so that

$$\begin{aligned}
\frac{\partial f}{\partial x_i} &= A_{ii}x_i + \frac{1}{2}\sum_{j,j\neq i} x_j(A_{ij} + A_{ji}) = \frac{1}{2}\sum_j x_j 2A_{ij} \\
&= \sum_j A_{ij}x_j = (A\boldsymbol{x})_i
\end{aligned}$$

This gives $\nabla f = A\boldsymbol{x}$. Finally we get

$$\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j}(\sum_j A_{ij}x_j) = A_{ij},$$

so that $\nabla^2 f = A$.

c. Show that, with f defined as in b., but with A not symmetric, we obtain that $\nabla f(\boldsymbol{x}) = \frac{1}{2}(A + A^T)\boldsymbol{x}$, and $\nabla^2 f = \frac{1}{2}(A + A^T)$. Verify that these formulas are compatibe with what you found in b. when A is symmetric.
Solution: As in b. we have that

$$f(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j} x_i A_{ij} x_j = \frac{1}{2}\sum_i A_{ii} x_i^2 + \frac{1}{2}\sum_{i,j,i\neq j} x_i A_{ij} x_j,$$

but the further simplifications now take the form

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= A_{ii} x_i + \frac{1}{2}\sum_{j,j\neq i} x_j (A_{ij} + A_{ji}) = \frac{1}{2}\sum_j x_j (A_{ij} + A_{ji}) \\ &= \frac{1}{2}\sum_j (A_{ij} + (A^T)_{ij}) x_j = (\frac{1}{2}(A + A^T)\boldsymbol{x})_i\end{aligned}$$

This gives $\nabla f = \frac{1}{2}(A + A^T)\boldsymbol{x}$. Finally we get

$$\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j}(\frac{1}{2}\sum_j (A_{ij} + (A^T)_{ij}) x_j) = \frac{1}{2}\sum_j (A_{ij} + (A^T)_{ij}),$$

so that $\nabla^2 f = \frac{1}{2}(A + A^T)$.

10. Consider $f(\boldsymbol{x}) = f(x_1, x_2) = x_1^2 + 3x_1 x_2 - 5x_2^2 + 3$. Determine the first order Taylor approximation to f at each of the points $(0,0)$ and $(2,1)$.
Solution: First note that $f(0,0) = 3$, and that $f(2,1) = 8$ We have that $\nabla f = (2x_1 + 3x_2, 3x_1 - 10x_2)$, and that $\nabla f(0,0) = (0,0)$, and $\nabla f(2,1) = (7,-4)$. The first order Taylor approximation at $(0,0)$ is thus

$$f(0,0) + \nabla f(0,0)^T(\boldsymbol{x} - (0,0)) = 3.$$

The first order Taylor approximation at $(2,1)$ is

$$\begin{aligned}f(2,1) + \nabla f(2,1)^T(\boldsymbol{x} - (2,1)) &= 8 + (7,-4)^T(x_1 - 2, x_2 - 1) \\ &= 8 + 7(x_1 - 2) - 4(x_2 - 1) = 7x_1 - 4x_2 - 2.\end{aligned}$$

11. Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}$. Show that A is positive definite. (Try to give two different proofs.)

12. Show that if A is positive definite, then its inverse is also positive definite.
Solution: If A is positive definite then its eigenvalues λ_i are positive. The eigenvalues of A^{-1} are $1/\lambda_i$, which also are positive, so that A^{-1} also is positive definite.

Chapter 2

1. We recall that $A \cap B$ consists of all points which lie both in A and B. Show that $A \cap B$ is convex when A and B are.

2. Suppose that f is a convex function defined on $\mathbb{R}$ which also is positive. Show that $g(x) = (f(x))^n$ also is convex.
Solution: Since f is convex we get that

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

But then we also have that

$$(f((1-\lambda)x + \lambda y))^n \leq ((1-\lambda)f(x) + \lambda f(y))^n.$$

Since $h(y) = y^n$ is convex for $y > 0$ (since $h''(y) > 0$), and since f is positive, we get that $((1-\lambda)f(x) + \lambda f(y))^n \leq (1-\lambda)(f(x))^n + \lambda(f(y))^n$, so that

$$\begin{aligned} g((1-\lambda)x + \lambda y) &= (f((1-\lambda)x + \lambda y))^n \\ &\leq (1-\lambda)(f(x))^n + \lambda(f(y))^n = (1-\lambda)g(x) + \lambda g(y). \end{aligned}$$

It follows that $g(x) = (f(x))^n$ is convex.

3. (Trial Exam UIO V2012) Assume that f, g are convex, positive, and increasing functions, both two times differentiable and defined on $\mathbb{R}$. Show that $h(x) = f(x)g(x)$ is convex.
Hint: Look at the second derivative of $h(x)$.
Solution: We have that $h'(x) = f'(x)g(x) + f(x)g'(x)$, and $h''(x) = f''(x)g(x) + f(x)g''(x) + 2f'(x)g'(x)$. Since f and g are convex we have that $f''(x) \geq 0$ and $g''(x) \geq 0$. Since the functions are increasing we have that $f'(x) \geq 0$ and $g'(x) \geq 0$. Since the functions also are positive we see that all three terms in the sum are ≥ 0 so that $h''(x) \geq 0$, and it follows that h also is convex.

4. Show that the previous result also holds for any f, g which are convex, positive, and increasing functions (i.e. they need not be differentiable).

5. (Exam UIO V2012)

a. Let f and g both be two times (continuously) differentiable functions defined on $\mathbb{R}$. Suppose also that f and g are convex, and that f is increasing. Show that $h(x) = f(g(x))$ is convex. This states that, in particular the function $f(\boldsymbol{x}) = e^{h(\boldsymbol{x})}$ (which we previsously just stated as convex without proof), is convex.
Hint: Compute the second derivative of $h(x)$, and consider its sign.
Solution: We have learnt that (continuously) differentiable functions are convex if and only if the Hessian is positive semidefinite (here this is translated to that the second derivative is ≥ 0). We have that $h'(x) = f'(g(x))g'(x)$, and that $h''(x) = f''(g(x))[g'(x)]^2 + f'(g(x))g''(x)$. Since f is convex we have that $f''(g(x)) \geq 0$. Since g also is convex we have that $g''(x) \geq 0$. Since f is increasing we also have that $f'(g(x)) \geq 0$. Therefore both terms in the sum must be ≥ 0, so that $h''(x) \geq 0$, so that h is convex.

b. Construct two convex functions f, g so that $h(x) = f(g(x))$ is not convex.

6. Let f be a convex function defined on $C \subset \mathbb{R}^n$. Show that $g(\boldsymbol{x}) = e^{f(\boldsymbol{x})}$ also is convex (i.e. the result from the previous exercise holds also when f is not differentiable).
Solution: Since f is convex we have that $f((1-\lambda)\boldsymbol{x}+\lambda\boldsymbol{y}) \leq (1-\lambda)f(\boldsymbol{x})+\lambda f(\boldsymbol{y})$. Since also $h(y) = e^y$ is an increasing function, we have that

$$g((1-\lambda)\boldsymbol{x}+\lambda\boldsymbol{y}) = e^{f((1-\lambda)\boldsymbol{x}+\lambda\boldsymbol{y})} \leq e^{(1-\lambda)f(\boldsymbol{x})+\lambda f(\boldsymbol{y})}.$$

Since $h(y) = e^y$ also is convex ($h''(y) = e^y > 0$) it follows that

$$g((1-\lambda)\boldsymbol{x}+\lambda\boldsymbol{y}) \leq e^{(1-\lambda)f(\boldsymbol{x})+\lambda f(\boldsymbol{y})} \leq (1-\lambda)e^{f(\boldsymbol{x})} + \lambda e^{f(\boldsymbol{y})} = (1-\lambda)g(\boldsymbol{x}) + \lambda g(\boldsymbol{y}),$$

and we have therefore shown that. If f was assumed to be two times differentiable, we could have done as follows: We have that

$$\begin{aligned}\nabla g(\boldsymbol{x}) &= h'(f(\boldsymbol{x}))\nabla f(\boldsymbol{x}) = e^{f(\boldsymbol{x})}\nabla f(\boldsymbol{x})\\ \nabla^2 g(\boldsymbol{x}) &= e^{f(\boldsymbol{x})}\nabla^2 f(\boldsymbol{x}) + e^{f(\boldsymbol{x})}\nabla f(\boldsymbol{x})(\nabla f(\boldsymbol{x}))^T.\end{aligned}$$

From this it is clear that $\nabla^2 g(\boldsymbol{x})$ is positive semidefinite: $\nabla^2 f(\boldsymbol{x})$ is positiv semidefinite since f is convex, and $\nabla f(\boldsymbol{x})(\nabla f(\boldsymbol{x}))^T$ is positive semidefinite since $\boldsymbol{h}^T\nabla f(\boldsymbol{x})(\nabla f(\boldsymbol{x}))^T\boldsymbol{h} = \|(\nabla f(\boldsymbol{x}))^T\boldsymbol{h}\|^2 \geq 0$ for all $\boldsymbol{h}$. Therefore, the sum $e^{f(\boldsymbol{x})}\nabla^2 f(\boldsymbol{x}) + e^{f(\boldsymbol{x})}\nabla f(\boldsymbol{x})(\nabla f(\boldsymbol{x}))^T$ is also positive semidefinite. It follows that $\nabla^2 g(\boldsymbol{x})$ is positive semidefinite, so that g is convex.

7. Let $S = \{(x, y, z) : z \geq x^2 + y^2\} \subset \mathbb{R}^3$. Sketch the set and verify that it is a convex set.
Solution: The function $f(x, y, z) = x^2 + y^2 - z$ is convex (the Hessian is positive semidefinite). The set in question can be written as the points where $f(x, y, z) \leq 0$, which is a sublevel set, and therefore convex.

8. Let $f : S \to \mathbb{R}$ be a differentiable function, where S is an open set in $\mathbb{R}$. Check that f is convex if and only if $f''(x) \geq 0$ for all $x \in S$.

9. Prove Proposition 2.3.

10. Prove Proposition 2.5.

11. Explain how you can write the LP problem max $\{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b},\ B\boldsymbol{x} = \boldsymbol{d},\ \boldsymbol{x} \geq \boldsymbol{0}\}$ as an LP problem of the form

$$\max\{\boldsymbol{c}^T\boldsymbol{x} : H\boldsymbol{x} \leq \boldsymbol{h},\ \boldsymbol{x} \geq \boldsymbol{0}\}$$

for suitable matrix H and vector $\boldsymbol{h}$.
Solution: Write B in row echelon form, to see which are pivot variables. Express these variables in terms of the free variables, and replace the pivot variables in all the equations. $A\boldsymbol{x} \geq \boldsymbol{b}$ then takes the form $C\boldsymbol{x} \geq \boldsymbol{b}$ (where $\boldsymbol{x}$ now is a shorter vector), and this can be written as $-C\boldsymbol{x} \leq -\boldsymbol{b}$, which is on the new form with $H = -C$, $\boldsymbol{h} = -\boldsymbol{b}$. Note that this strategy rewrites the vector $\boldsymbol{c}$ to a shorter vector.

12. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_t \in \mathbb{R}^n$ and let C be the set of vectors of the form

$$\sum_{j=1}^{t} \lambda_j \boldsymbol{x}_j$$

where $\lambda_j \geq 0$ for each $j = 1, \dots, t$, and $\sum_{j=1}^{t} \lambda_j = 1$. Show that C is convex. Make a sketch of such a set in $\mathbb{R}^3$.
Solution: Let $\boldsymbol{y} = \sum_{j=1}^{t} \lambda_j \boldsymbol{x}_j$ and $\boldsymbol{z} = \sum_{j=1}^{t} \mu_j \boldsymbol{x}_j$, where all $\lambda_j, \mu_j \geq 0$, and $\sum_{j=1}^{t} \lambda_j = 1$, $\sum_{j=1}^{t} \mu_j = 1$. For any $0 \leq \lambda \leq 1$ we have that

$$(1-\lambda)\boldsymbol{y} + \lambda\boldsymbol{z} = (1-\lambda)\sum_{j=1}^{t} \lambda_j \boldsymbol{x}_j + \lambda \sum_{j=1}^{t} \mu_j \boldsymbol{x}_j = \sum_{j=1}^{t} ((1-\lambda)\lambda_j + \lambda\mu_j)\boldsymbol{x}_j.$$

The sum of the coefficients here is

$$\sum_{j=1}^{t} ((1-\lambda)\lambda_j + \lambda\mu_j) = (1-\lambda)\sum_{j=1}^{t} \lambda_j + \lambda \sum_{j=1}^{t} \mu_j = 1 - \lambda + \lambda = 1,$$

so that C is a convex set.

13. Show that $f(\boldsymbol{x}) = e^{\sum_{j=1}^{n} x_j}$ is a convex function.
Solution: Follows from Proposition 2.3, since $f(x) = e^x$ is convex, and $\boldsymbol{H}(\boldsymbol{x}) = \sum_{j=1}^{n} x_j$ is affine.

14. Assume that f and g are convex functions defined on an interval I. Which of the following functions are convex or concave?

a. λf where $\lambda \in \mathbb{R}$,
Solution: λf is convex if $\lambda \geq 0$, concave if $\lambda \leq 0$.

b. $\min\{f, g\}$,
Solution: $\min\{f, g\}$ may be neither convex or concave, consider the functions $f(x) = x^2$, $g(x) = (x-1)^2$.

c. $|f|$.

Solution: $|f|$ may be neither convex or concave, consider the function $f(x) = x^2 - 1$.

15. Let $f : [a, b] \to \mathbb{R}$ be a convex function. Show that

$$\max\{f(x) : x \in [a, b]\} = \max\{f(a), f(b)\}$$

i.e., a convex function defined on closed real interval attains its maximum in one of the endpoints.

16. (Trial Exam UIO V2012) Show that $\max\{f, g\}$ is a convex function when f and g are convex (we define $\max\{f, g\}$ by $\max\{f, g\}(\boldsymbol{x}) = \max(f(\boldsymbol{x}), g(\boldsymbol{x}))$)).

Solution: That f is convex means that $f((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}) \le (1-\lambda)f(\boldsymbol{x}) + \lambda f(\boldsymbol{y})$ for all λ between 0 and 1. We have that

$$\begin{aligned} f((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}) &\le (1-\lambda)f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) \le (1-\lambda)\max\{f, g\}(\boldsymbol{x}) + \lambda\max\{f, g\}(\boldsymbol{y}) \\ g((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}) &\le (1-\lambda)g(\boldsymbol{x}) + \lambda g(\boldsymbol{y}) \le (1-\lambda)\max\{f, g\}(\boldsymbol{x}) + \lambda\max\{f, g\}(\boldsymbol{y}), \end{aligned}$$

but then also

$$\max\{f, g\}((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{y}) \le (1-\lambda)\max\{f, g\}(\boldsymbol{x}) + \lambda\max\{f, g\}(\boldsymbol{y}),$$

so that $\max\{f, g\}$ also is convex.

17. Let $f : \langle 0, \infty\rangle \to \mathbb{R}$ and define the function $g : \langle 0, \infty\rangle \to \mathbb{R}$ by $g(x) = xf(1/x)$. Why is the function $x \to xe^{1/x}$ convex?

18. Let $C \subseteq \mathbb{R}^n$ be a convex set and consider the distance function d_C defined by $d_C(x) = \inf\{\|x - y\| : y \in C\}$. Show that d_C is a convex function.

Chapter 3

1. Show that the problem of solving nonlinear equations (3.1) may be transformed into a nonlinear optimization problem. (Hint: Square each component function and sum these up!)
Solution: $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$ is equivalent to $\|\boldsymbol{F}(\boldsymbol{x})\|^2 = \sum_i F_i(\boldsymbol{x}) = 0$, where F_i are the component functions of $\boldsymbol{F}$. Solving $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$ thus is equivalent to showing that 0 is the minimum value of $\sum_i F_i(\boldsymbol{x})$.

2. Let $T : \mathbb{R} \to \mathbb{R}$ be given by $T(x) = (3/2)(x - x^3)$. Draw the graph of this function, and determine its fixed points. Let x^* denote the largest fixed point. Find, using your graph, an interval I containing x^* such that the fixed point algorithm with an initial point in I will guaranteed converge towards x^*. Then try the fixed point algorithm with starting point $x_0 = \sqrt{5/3}$.
Solution: Here we construct the function $f(x) = T(x) - x = x/2 - 3x^3/2$, which has derivative $f'(x) = 1/2 - 9x^2/2$. We can then run Newton's method as follows:

```
newtonmult(sqrt(5/3),@(x)(0.5*x-1.5*x^3),@(x)(0.5-4.5*x^2))
```

This converges to the zero we are looking for, which we easily compute as $x = \sqrt{1/3}$.

3. Let $\alpha \in \mathbb{R}_+$ be fixed, and consider $f(x) = x^2 - \alpha$. Then the zeros are $\pm\sqrt{\alpha}$. Write down the Newton's iteration for this problem. Let $\alpha = 2$ and compute the first three iterates in Newton's method when $x_0 = 1$.

4. For any vector norm $\|\cdot\|$ on $\mathbb{R}^n$, we can more generally define a corresponding *operator norm* for $n \times n$ matrices by

$$\|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

a. Explain why this supremum is attained.
Solution: The function $x \to \|A\boldsymbol{x}\|$ is continuous, and any continuous function achieves a supremum in a closed set (here $\|\boldsymbol{x}\| = 1$).

In the rest of this exercise we will use the vector norm $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_1 = \sum_{j=1}^n |x_j|$ on $\mathbb{R}^n$.

b. For $n=2$, draw the sublevel set $\{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_1 \le 1\}$.
Solution: For $n=2$, it is clear that the sublevel set is the square with corners $(1,0)$, $(-1,0)$, $(0,1)$, $(0,-1)$.

c. Show that $f(\boldsymbol{x}) = \|A\boldsymbol{x}\|$ is convex for any n, and show that the maximum of f on the set $\{\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$ is attained in a point $\boldsymbol{x}$ on the form $\pm\boldsymbol{e}_k$.
Hint: For the second statement, use Jensen's inequality with $\boldsymbol{x}^j = \pm\boldsymbol{e}_j$ (Theorem 2.4).
Solution: The function $f(\boldsymbol{x}) = \|A\boldsymbol{x}\|$ is the composition of a convex function and an affine function, so that it must be convex. If $\boldsymbol{x} \in \mathbb{R}^n$ and $\|\boldsymbol{x}\|_1 = 1$, we can write $\boldsymbol{x} = \sum_{i=1}^n \lambda_i \boldsymbol{v}_i$, where $0 \le \lambda_i \le 1$, $\sum_{i=1}^n \lambda_i = 1$, and $\boldsymbol{v}_i = \pm\boldsymbol{e}_i$ (i.e. it absorbs the sign of the ith component). If $\boldsymbol{w}$ is the vector among $\{\pm\boldsymbol{e}_j\}_j$ so that $f(\pm\boldsymbol{e}_j) \le f(\boldsymbol{w})$ for all j and all signs, Jensen's inequality (Theorem 2.4) gives

$$f(\boldsymbol{x}) = f\left(\sum_{i=1}^n \lambda_i \boldsymbol{v}_i\right) \le \sum_{i=1}^n \lambda_i f(\boldsymbol{v}_i) \le \sum_{i=1}^n \lambda_i f(\boldsymbol{w}) = f(\boldsymbol{w}),$$

so that f assumes its maximum in $\boldsymbol{w}$.

d. Show that, for any $n \times n$-matrix A, $\|A\| = \sup_k \sum_{i=1}^n |a_{ik}|$, where a_{ij} are the entries of A (i.e. the biggest sum of absolute values in a column).
Solution: Since the supremum is attained for some $\boldsymbol{w} = \pm\boldsymbol{e}_k$, the maximum is

$$\|A\boldsymbol{w}\|_1 = \|\pm \mathrm{col}_k A\| = \sum_{i=1}^n |\pm a_{ik}| = \sum_{i=1}^n |a_{ik}|.$$

It is now clear that $\|A\| = \sup_k \sum_{i=1}^n |a_{ik}|$.

5. Consider a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\boldsymbol{x}) = A\boldsymbol{x}$ where A is an $n \times n$ matrix. When is T a contraction w.r.t. the vector norm $\|\cdot\|_1$?
Solution: We are asked to find for which A we have that $\|A\boldsymbol{x}\|_1 < \|\boldsymbol{x}\|_1$ for any $\boldsymbol{x}$. From the previous exercise we know that this happens if and only if $\|A\| < 1$, i.e. when $\sum_{i=1}^n |a_{ik}| < 1$ for all k.

6. Test the function `newtonmult` on the equations given initially in Section 3.1.
Solution: You can write

```
newtonmult(x0,...
  @(x)([x(1)^2-x(1)/x(2)^3+cos(x(1))-1; 5*x(1)^4+2*x(1)^3-tan(x(1)*x(2)^8)-3]),...
  @(x)([2*x(1)-1/x(2)^3-sin(x(1))                              3*x(1)/x(2)^4; ...
        20*x(1)^3+6*x(1)^2-x(2)^8/(cos(x(1)*x(2)^8))^2      -8*x(1)*x(2)^7/(cos(x(1)*x(2)^8))^
)
```

7. In this exercise we will implement Broyden's method with Matlab.

a. Given a value $\boldsymbol{x}_0$, implement a function which computes an estimate of $\boldsymbol{F}'(\boldsymbol{x}_0)$ by estimating the partial derivatives of $\boldsymbol{F}$, using a numerical differentiation method and step size of you own choosing.

b. Implement a function

```
function x=broyden(x0,F)
```

which returns an estimate of a zero of $\boldsymbol{F}$ using Broyden's method. Your method should set B_0 to be the matrix obtained from the function in a. Just indicate where line search along the search direction should be performed in your function, without implementing it. The function should work as `newtonmult` in that it terminates after a given number of iterations, or after precision of a given accuracy has been obtained.

Chapter 4

1. Consider the function $f(x_1, x_2) = x_1^2 + ax_2^2$ where $a > 0$ is a parameter. Draw some of the level sets of f (for different levels) for each a in the set $\{1, 4, 100\}$. Also draw the gradient in a few points on these level sets.

2. State and prove a theorem similar to Theorem 4.1 for maximization problems.

3. Let $f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ where A is a symmetric $n \times n$ matrix. Assume that A is indefinite, so it has both positive and negative eigenvalues. Show that $\boldsymbol{x} = \boldsymbol{0}$ is a saddlepoint of f.

4. Let $f(x_1, x_2) = 4x_1 + 6x_2 + x_1^2 + 2x_2^2$. Find all stationary points and determine if they are minimum, maximum or saddlepoints. Do the same for the function $g(x_1, x_2) = 4x_1 + 6x_2 + x_1^2 - 2x_2^2$.

Solution: The gradient of f is $\nabla f = (4 + 2x_1, 6 + 4x_2)$, and the Hessian matrix is $\nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, which is positive definite. The only stationary point is $(-2, -3/2)$, which is a minimum.

The gradient of g is $\nabla g = (4 + 2x_1, 6 - 4x_2)$, and the Hessian matrix is $\nabla^2 g = \begin{pmatrix} 2 & 0 \\ 0 & -4 \end{pmatrix}$, which is indefinite. The only stationary point is $(-2, 3/2)$, which must be a saddle point.

5. Let the function f be given by $f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2 + 1$.

a. Compute the search direction $\boldsymbol{d}_k$ which is chosen by the steepest descent method in the point $\boldsymbol{x}_k = (2, 3)$.

b. Compute in the same way the search direction $\boldsymbol{d}_k$ which is chosen when we instead use Newton's method in the point $\boldsymbol{x}_k = (2, 3)$.

6. The function $f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ is called the *Rosenbrock function*. Compute the gradient and the Hessian matrix at every point $\boldsymbol{x}$. Find every local minimum. Also draw some of the level sets (contour lines) of f using Matlab.

Solution: The gradient is $\nabla f = (-400x_1(x_2-x_1^2)-2(1-x_1), 200(x_2-x_1^2))$. The Hessian matrix is

$$\nabla^2 f = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}.$$

Clearly the only stationary point is $x = (1,1)$, and we get that

$$\nabla^2 f(1,1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}.$$

It is straightforward to check that this matrix is positive definite, so that $(1,1)$ is a local minimum.

7. Let $f(\boldsymbol{x}) = (1/2)\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$ where A is a positive definite $n\times n$ matrix. Consider the steepest descent method applied to the minimization of f, where we assume exact line search is used. Assume that the search direction happens to be equal to an eigenvector of A. Show that then the minimum is reached in just one step.
Solution: The steepest descent method takes the form

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k),$$

where $\nabla f(\boldsymbol{x}_k) = A\boldsymbol{x}_k - \boldsymbol{b}$. We have that

$$\begin{aligned}
f(\boldsymbol{x}_{k+1}) &= (1/2)\boldsymbol{x}_{k+1}^T A\boldsymbol{x}_{k+1} - \boldsymbol{b}^T\boldsymbol{x}_{k+1} \\
&= \frac{1}{2}\left(\nabla f(\boldsymbol{x}_k)^T A\nabla f(\boldsymbol{x}_k)\right)\alpha_k^2 - \frac{1}{2}\left(\boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k\right)\alpha_k \\
&+ \frac{1}{2}\boldsymbol{x}_k^T A\boldsymbol{x}_k - \boldsymbol{b}^T(\boldsymbol{x}_k - \alpha_k\nabla f(\boldsymbol{x}_k)) \\
&= \frac{1}{2}\left(\nabla f(\boldsymbol{x}_k)^T A\nabla f(\boldsymbol{x}_k)\right)\alpha_k^2 \\
&+ \left(\boldsymbol{b}^T\nabla f(\boldsymbol{x}_k) - \frac{1}{2}\left(\boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k\right)\right)\alpha_k \\
&+ \frac{1}{2}\boldsymbol{x}_k^T A\boldsymbol{x}_k - \boldsymbol{b}^T\boldsymbol{x}_k.
\end{aligned}$$

Now, since $\nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k$ is a scalar, it is in particular symmetric, so that

$$\nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k = (\nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k)^T = \boldsymbol{x}_k^T A^T\nabla f(\boldsymbol{x}_k) = \boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k),$$

where we have used that A is symmetric. We conclude that $\nabla f(\boldsymbol{x}_k)^T A\boldsymbol{x}_k = \boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k)$. We can thus simplify what we found above to

$$\begin{aligned}
&f(\boldsymbol{x}_{k+1}) \\
&= \frac{1}{2}\left(\nabla f(\boldsymbol{x}_k)^T A\nabla f(\boldsymbol{x}_k)\right)\alpha_k^2 + \left(\boldsymbol{b}^T\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k)\right)\alpha_k + \frac{1}{2}\boldsymbol{x}_k^T A\boldsymbol{x}_k - \boldsymbol{b}^T\boldsymbol{x}_k \\
&= \frac{1}{2}\lambda\|\nabla f(\boldsymbol{x}_k)\|^2\alpha_k^2 + \left(\boldsymbol{b}^T\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k)\right)\alpha_k + \frac{1}{2}\boldsymbol{x}_k^T A\boldsymbol{x}_k - \boldsymbol{b}^T\boldsymbol{x}_k,
\end{aligned}$$

where we also have used that $\nabla f(\boldsymbol{x}_k)$ is an eigenvector of A, that A is symmetric, so that $A\nabla f(\boldsymbol{x}_k) = \lambda\nabla f(\boldsymbol{x}_k)$ and $\nabla f(\boldsymbol{x}_k)^T A = \lambda\nabla f(\boldsymbol{x}_k)^T$, where λ is the corresponding eigenvalue. If we are to apply exact line search, we need to minimize this expression w.r.t. α_k. This can be done by taking the derivative w.r.t. α_k and setting this to 0. If we do this we get

$$\begin{aligned}\alpha_k &= -\frac{\boldsymbol{b}^T\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}_k^T A\nabla f(\boldsymbol{x}_k)}{\lambda\|\nabla f(\boldsymbol{x}_k)\|^2} = \frac{\left(\boldsymbol{x}_k^T A - \boldsymbol{b}^T\right)\nabla f(\boldsymbol{x}_k)}{\lambda\|\nabla f(\boldsymbol{x}_k)\|^2} \\ &= \frac{(A\boldsymbol{x}_k - \boldsymbol{b})^T\nabla f(\boldsymbol{x}_k)}{\lambda\|\nabla f(\boldsymbol{x}_k)\|^2} = \frac{\nabla f(\boldsymbol{x})^T\nabla f(\boldsymbol{x}_k)}{\lambda\|\nabla f(\boldsymbol{x}_k)\|^2} = \frac{1}{\lambda}.\end{aligned}$$

This means that $\alpha_k = \frac{1}{\lambda}$ is the step size we should use when we perform exact line search. We now compute that

$$\begin{aligned}\nabla f(\boldsymbol{x}_{k+1}) &= A\boldsymbol{x}_{k+1} - \boldsymbol{b} = A\left(\boldsymbol{x}_k - \frac{1}{\lambda}\nabla f(\boldsymbol{x}_k)\right) - \boldsymbol{b} \\ &= A\boldsymbol{x}_k - \frac{1}{\lambda}A\nabla f(\boldsymbol{x}_k) - \boldsymbol{b} = A\boldsymbol{x}_k - \nabla f(\boldsymbol{x}_k) - \boldsymbol{b} \\ &= \nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_k) = \boldsymbol{0},\end{aligned}$$

which shows that the minimum is reached in one step.

8. Consider the second order Taylor approximation

$$T_f^2(\boldsymbol{x};\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T\boldsymbol{h} + (1/2)\boldsymbol{h}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{h}.$$

a. Show that $\nabla_{\boldsymbol{h}} T_f^2 = \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{h}$.
Solution: This is simply Exercise 9 in Chapter1.3.

b. Minimizing T_f^2 with respect to $\boldsymbol{h}$ implies solving $\nabla_{\boldsymbol{h}} T_f^2 = \boldsymbol{0}$, i.e. $\nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{h} = \boldsymbol{0}$ from a.. If $\nabla^2 f(\boldsymbol{x})$ is positive definite, explain that it also is invertible, so that this equation has the unique solution $\boldsymbol{h} = -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$, as previously noted for the Newton step.
Solution: If $\nabla^2 f(\boldsymbol{x})$ is positive definite, its eigenvalues are positive, so that the determinant is positive, and that the matrix is invertible. $\boldsymbol{h} = -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$ follows after multiplying with the inverse.

9. We want to find the minimum of $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$, defined on $\mathbb{R}^n$. Formulate one step with Newton's method, and one step with the steepest descent method, where you set the step size to $\alpha_k = 1$. Which of these methods works best for finding the minimum for functions on this form?
Solution: We have that $\nabla f(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$, and that $\nabla^2 f(\boldsymbol{x}) = A$. One step with Newton's method with $\alpha_k = 1$ is therefore

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}))^{-1}\nabla f(\boldsymbol{x}) = \boldsymbol{x}_k - A^{-1}(A\boldsymbol{x}_k - \boldsymbol{b}) = A^{-1}\boldsymbol{b}.$$

One step with the steepest descent method with $\alpha_k = 1$ becomes

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \nabla f(\boldsymbol{x}_k) = \boldsymbol{x}_k - (A\boldsymbol{x}_k - \boldsymbol{b}) = (I - A)\boldsymbol{x}_k + \boldsymbol{b}.$$

It is clear that Newton's method is best here, since this corresponds to finding the minimum for the second order approximation where $\alpha_k = 1$, and here the function is equal to the second order approximation.

10. Implement the steepest descent method. Test the algorithm on the functions in exercises 4 and 6. Use different starting points.
Solution: Here we have said nothing about the step length, but we can implement this as in the function `newtonbacktrack` as follows:

```
function [x,numit]=steepestdescent(f,df,x0)
  epsilon=10^(-3);
  x=x0;
  maxit=100;
  for numit=1:maxit
      d=-df(x);
      eta=-df(x)'*d;
      if eta/2<epsilon
          break;
      end
      % Armijos rule
      beta=0.2; s=0.5; sigma=10^(-3);
      m=0;
      while (f(x)-f(x+beta^m*s*d) < -sigma *beta^m*s *(df(x))'*d)
          m=m+1;
      end
      alpha = beta^m*s;

      x=x+alpha*d;
  end
```

The algorithm can be tested on the first function from Exercise 4 as follows:

```
f=@(x)(4*x(1)+6*x(2)+x(1)^2+2*x(2)^2);
df=@(x)([4+2*x(1);6+4*x(2)])
steepestdescent(f,df,[-1;-1])
```

11. What can go wrong when you apply backtracking line search (Equation (4.7)) to a function f where $\nabla^2 f$ er negative definite (i.e. all eigenvalues of $\nabla^2 f$ are negative)? Hint: Substitute the Taylor approximation

$$f(\boldsymbol{x}_k + \beta^m s\boldsymbol{d}_k) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T(\beta^m s\boldsymbol{d}_k)$$

in Equation (4.7), and remember that σ there is chosen so that $\sigma < 1$.

12. Write a function `newtonbacktrack` which performs Newton's method for unconstrained optimization. The input parameters are the function, its gradient, its Hesse matrix, and the initial point. The function should also return the number of iterations, and at each iteration write the corresponding function value. Use backtracking line search to compute the step size, i.e. compute m_k from Equation (4.7) with $\beta = 0.2$, $s = 0.5$, $\sigma = 10^{-3}$, and use $\alpha = \beta^{m_k} s$ as the step size. Test the algorithm on the functions in exercises 4 and 6. Use different starting points.
Solution: The function can be implemented as follows:

```
function [x,numit]=newtonbacktrack(f,df,d2f,x0)
  beta=0.2; s=0.5; sigma=10^(-3);
  epsilon=10^(-3);
  x=x0;
  maxit=100;
  for numit=1:maxit
      d=-d2f(x)\df(x);
      eta=-df(x)'*d;
      if eta/2<epsilon
          break;
      end
      m=0;
      % Armijos rule
      while (f(x)-f(x+beta^m*s*d) < -sigma *beta^m*s *(df(x))'*d)
        m=m+1;
      end
      x=x+beta^m*s*d;
  end
```

13. Let us return to the maximum likelihood example 1.3.

a. Run the function `newtonbacktrack` with parameters being the function f and its and derivaties defined as in Example 1.3 with $n = 10$ and

$$\mathbf{x} = (0.4992, -0.8661, 0.7916, 0.9107, 0.5357, 0.6574, 0.6353, 0.0342, 0.4988, -0.4607)$$

Use the start value $\alpha_0 = 0$ for Newtons method. What estimate for the minimum of f (and thereby α) did you obtain?

b. The ten measurements from a. were generated from a probability distribution where $\alpha = 0.5$. The answer you obtained was quite far from this. Let us therefore take a look at how many measurements we should use in order to get quite precise estimates for α. You can use the function

```
function ret=randmuon(alpha,m,n)
```

to generate an $m \times n$-matrix with measurements generated with a probability distribution with a given parameter α. This function can be found at the homepage of the book.

With $\alpha = 0.5$, generate $n = 10$ measurements with the help of the function `randmuon`, and find the maximum likelihood estimate as above. Repeat this 10 times, and plot the ten estimates you obtain. Repeat for $n = 1000$, and for $n = 100000$ (in all cases you are supposed to plot 10 maximum likelihood estimates). How many measurements do we need in order to obtain maximum likelihood estimates which are reliable?

Note that it is possible for the maximum likelihood estimates you obtain here to be outside the domain of definition $[-1, 1]$. You need not take this into account.

Chapter 5

1. In the plane consider a rectangle R with sides of length x and y and with perimeter equal to α (so $2x+2y=\alpha$). Determine x and y so that the area of R is largest possible.

2. Consider the optimization problem

$$\text{minimize } f(x_1,x_2) \text{ subject to } (x_, x_2) \in C$$

where $C = \{(x_1,x_2) \in \mathbb{R}^2 : x_1, x_2 \geq 0, 4x_1 + x_2 \geq 8, 2x_1 + 3x_3 \leq 12\}$. Draw the feasible set C in the plane. Find the set of optimal solutions in each of the cases given below.

a. $f(x_1,x_2) = 1$.

b. $f(x_1,x_2) = x_1$.

c. $f(x_1,x_2) = 3x_1 + x_2$.

d. $f(x_1,x_2) = (x_1-1)^2 + (x_2-1)^2$.

e. $f(x_1,x_2) = (x_1-10)^2 + (x_2-8)^2$.

3. Solve

$$\max\{x_1 x_2 \cdots x_n : \sum_{j=1}^{n} x_j = 1, x_j \geq 0\}.$$

Solution: This is the same as finding the minimum of $f(x_1,\ldots,x_n) = -x_1 x_2 \cdots x_n$. This boils down to the equations $-\prod_{i\neq j} x_i = 1$, since clearly the minimum is not attained when there are any active constraints. This implies that $x_1 = \ldots = x_n$, so that all $x_i = 1/n$. It is better to give a direct argument here that this must be a minimum, than to attempt to analyse the second order conditions for a minimum.

4. Let $S = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\| = 1\}$ be the unit circle in the plane. Let $\boldsymbol{a} \in \mathbb{R}^2$ be a given point. Formulate the problem of finding a nearest point in S to $\boldsymbol{a}$ as a nonlinear optimization problem. How can you solve this problem directly using a geometrical argument?

Solution: We can formulate the problem as finding the minimum of $f(x_1,x_2) = (x_1-a_1)^2+(x_2-a_2)^2$ subject to the constraint $h_1(x_1,x_2) = x_1^2+x_2^2 = 1$. The minimum can be found geometrically by drawing a line which passes through $\boldsymbol{a}$ and the origin, and reading the intersection with the unit circle. This follows also from that ∇f is parallel to $\boldsymbol{x}-\boldsymbol{a}$, ∇h_1 is parallel to $\boldsymbol{x}$, and from that the KKT-conditions say that these should be parallel.

5. Let S be the unit circle from the previous exercise. Let $\boldsymbol{a}_1,\boldsymbol{a}_2$ be two given points in the plane. Let $f(\boldsymbol{x}) = \sum_{i=1}^2 \|\boldsymbol{x}-\boldsymbol{a}_i\|^2$. Formulate this as an optimization problem and find its Lagrangian function L. Find the stationary points of L, and use this to solve the optimization problem.

6. Solve

$$\text{minimize } x_1+x_2 \text{ subject to } x_1^2+x_2^2 = 1.$$

using the Lagrangian, see Theorem 5.1. Next, solve the problem by eliminating x_2 (using the constraint).

Solution: We rewrite the constraint as $g_1(x_1,x_2) = x_1^2+x_2^2-1 = 0$, and get that $\nabla g_1(x_1,x_2) = (2x_1,2x_2)$. Clearly all points are regular, since $\nabla g_1(x_1,x_2) \neq 0$ whenever $g_1(x_1,x_2) = 0$. Since $\nabla f = (1,1)$ we get that the gradient of the Lagrangian is

$$\begin{pmatrix}1\\1\end{pmatrix} + \lambda\begin{pmatrix}2x_1\\2x_2\end{pmatrix} = \mathbf{0},$$

which gives that $x_1 = x_2$. This gives us the two possible feasible points $(1/\sqrt{2},1/\sqrt{2})$ and $(-1/\sqrt{2},-1/\sqrt{2})$. For the first we see that $\lambda = -1/\sqrt{2}$, for the second we see that $\lambda = 1/\sqrt{2}$. The Hessian of the Lagrangian is $\lambda\begin{pmatrix}2&0\\0&2\end{pmatrix}$. For the point $(1/\sqrt{2},1/\sqrt{2})$ this is negative definite since λ is negative, for the point $(-1/\sqrt{2},-1/\sqrt{2})$ this is positive definite since λ is positive. From the second order conditions it follows that the minimum is attained in $(-1/\sqrt{2},-1/\sqrt{2})$.

If we instead eliminated x_2 we must write $x_2 = -\sqrt{1-x_1^2}$ (since the positive square root gives a bigger value for f), so that we must minimize $f(x) = x-\sqrt{1-x^2}$ subject to the constraint $-1 \le x \le 1$. The derivative of this is $1+\frac{x}{\sqrt{1-x^2}}$, which is zero when $x = -\frac{1}{\sqrt{2}}$, which we found above. We also could have found this by considering the two inequality constraints $-x-1 \le 0$ and $x-1 \le 0$.

If the first one of these is active (i.e. $x = -1$), the KKT conditions say that $f'(-1) > 0$. However, this is not the case since $f'(x) \to -\infty$ when $x \to -1_+$. If the second constraint is active (i.e. $x = 1$), the KKT conditions say that $f'(1) < 0$. This is not the case since $f'(x) \to \infty$ when $x \to 1-$. When we have no active constraint, the problem boils down to setting the derivative to zero, in which case we get the solution we already have found.

7. Let $g(x_1,x_2) = 3x_1^2+10x_1x_2+3x_2^2-2$. Solve

$$\min\{\|(x_1,x_2)\| : g(x_1,x_2) = 0\}.$$

8. Same question as in previous exercise, but with $g(x_1, x_2) = 5x_1^2 - 4x_1x_2 + 4x_2^2 - 6$.

9. Let f be a two times differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Consider the optimization problem

$$\text{minimize } f(\boldsymbol{x}) \text{ subject to } x_1 + x_2 + \cdots + x_n = 1.$$

Characterize the stationary points (find the equation they satisfy).
Solution: We define $h_1(x_1, \ldots, x_n) = x_1 + \ldots + x_n - 1$, and find that $\nabla h_1 = (1, 1, \ldots, 1)$. The stationary points are characterized by $\nabla f + \boldsymbol{\lambda}^T (1, 1, \ldots, 1) = \mathbf{0}$, which has a solution exactly when $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \ldots = \frac{\partial f}{\partial x_n}$.

10. Consider the previous exercise. Explain how to convert this into an unconstrained problem by eliminating x_n.
Solution: We substitute $x_n = 1 - x_1 - \ldots - x_{n-1}$ in the expression for f, to turn the problem into one of minimizing a function in $n-1$ variables.

11. Let A be a real symmetric $n \times n$ matrix. Consider the optimization problem

$$\max\left\{\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} : \|\boldsymbol{x}\| = 1\right\}$$

Rewrite the constraint as $\|\boldsymbol{x}\| - 1 = 0$ and show that an optimal solution of this problem must be an eigenvector of A. What can you say about the Lagrangian multiplier?
Solution: The problem can be rewritten to the following minimation problem:

$$\min\left\{-\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} : h_1(\boldsymbol{x}) = \|\boldsymbol{x}\| - 1 = 0\right\}.$$

We have that $\nabla f(\boldsymbol{x}) = -A\boldsymbol{x}$, and $\nabla h_1(\boldsymbol{x}) = \frac{2\boldsymbol{x}}{2\|\boldsymbol{x}\|} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$. Clearly all points are regular, and we get that

$$\nabla f + \lambda \nabla h_1 = -A\boldsymbol{x} + \lambda \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \mathbf{0}.$$

Since we require that $\|\boldsymbol{x}\| = 1$ we get that $A\boldsymbol{x} = \lambda\boldsymbol{x}$. In other words, the optimal point $\boldsymbol{x}$ is an eigenvector of A, and the Lagrange multiplier is the corresponding eigenvalue.

12. Solve

$$\min\{(1/2)(x_1^2 + x_2^2 + x_3^2) : x_1 + x_2 + x_3 \leq -6\}.$$

Solution: Define $f(x_1, x_2, x_3) = (1/2)(x_1^2 + x_2^2 + x_3^2)$ and $g_1(x_1, x_2, x_3) = x_1 + x_2 + x_3$. We have that $\nabla f = (x_1, x_2, x_3)$, $\nabla g_1 = (1, 1, 1)$. Clearly all points are regular points. If there are no active constraints, we must have that $\nabla f = \mathbf{0}$, so that $x_1 = x_2 = x_3 = 0$, which does not fulfill the constraint. If the constraint is active we must have that $(x_1, x_2, x_3) + \mu(1, 1, 1) = \mathbf{0}$ for some $\mu \leq 0$, which is satisfied when $x_1 = x_2 = x_3 < 0$. Clearly we must have that $x_1 = x_2 = x_3 = -2$. The Hessian of $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is easily computed to be positive definite, so that we have found a minimum.

13. Solve
$$\min\{(x_1-3)^2+(x_2-5)^2+x_1x_2 : 0 \le x_1, x_2 \le 1\}.$$

Solution: We need to minimize $f(x_1,x_2)=(x_1-3)^2+(x_2-5)^2+x_1x_2$ subject to the constraints

$$\begin{aligned}
g_1(x_1,x_2) &= -x_1 \le 0\\
g_2(x_1,x_2) &= -x_2 \le 0\\
g_3(x_1,x_2) &= x_1-1 \le 0\\
g_4(x_1,x_2) &= x_2-1 \le 0.
\end{aligned}$$

We have that $\nabla f = (2(x_1-3)+x_2, 2(x_2-5)+x_1)$, and $\nabla g_1 = (-1,0)$, $\nabla g_2 = (0,-1)$, $\nabla g_3 = (1,0)$, $\nabla g_4 = (0,1)$.

Clearly all points are regular: even though the gradients ∇g_1 and ∇g_3 are linearly dependent, they can't be active at the same time, and similarly for g_2 and g_4. Also any of $\nabla g_1, \nabla g_3$ is independent from any of $\nabla g_2, \nabla g_4$, so that linear dependence between active gradients is impossible, so that all points are regular.

The KKT conditions are $\nabla f + \mu_1\nabla g_1 + \mu_2\nabla g_2 + \mu_3\nabla g_3 + \mu_4\nabla g_4 = \mathbf{0}$, where $\mu_i = 0$ if g_i is not active. This can be written

$$\begin{pmatrix} 2(x_1-3)+x_2 \\ 2(x_2-5)+x_1 \end{pmatrix} + \begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

subject to the constraints $0 \le x_1, x_2 \le 1$, $\mu_i \ge 0$. Here we have grouped together the gradients with negative signs on the right hand side. Note first that, due to the constraints $0 \le x_1, x_2 \le 1$, the entries in $\nabla f = (2(x_1-3)+x_2, 2(x_2-5)+x_1)$ are both negative. Since all $\mu_i \ge 0$, it is impossible for any of μ_3 and μ_4 to be zero, since this then would imply that one of μ_1 or μ_2 is negative, by computing the left hand side above. But this implies that both g_3 and g_4 must be active, so that $x_1 = x_2 = 1$. Then clearly g_1 and g_2 are not active, so that $\mu_1 = \mu_2 = 0$, and we get the equation (by inserting $x_1 = x_2 = 1$)

$$\begin{pmatrix} -3 \\ -7 \end{pmatrix} + \begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

so that $\mu_3 = 3$, $\mu_4 = 7$, which is allowed. This means that $(1,1)$ is the only candidate for minimum, and the minimum value is $f(1,1) = 21$

14. Solve
$$\min\{x_1 + x_2 : x_1^2 + x_2^2 \le 2\}.$$

Solution: We can define $g_1(x_1,x_2) = x_1^2+x_2^2-2$, so that the only constraint is $g_1(x_1,x_2) \le 0$. We have that $\nabla g_1 = (2x_1, 2x_2)$, and this can be zero if and only if $x_1 = x_2 = 0$. However $g_1(0,0) = -2 < 0$, so that the equality is not active. This means that all points are regular for this problem.

We compute that $\nabla f = (1,1)$. If g_1 is not an active inequality, the KKT conditions say that $\nabla f = 0$, which is impossible. If g_1 is active, we get that

$$\nabla f(x_1, x_2) + \mu \nabla g_1(x_1, x_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

so that $1 = -2\mu x_1$ and $1 = -2\mu x_2$ for some $\mu \geq 0$. This is satisfied if $x_1 = x_2$ is negative. For g_1 to be active we must have that $x_1^2 + x_2^2 = 2$, which implies that $x_1 = x_2 = -1$. We have that $f(-1,-1) = -2$.

15. Write down the KKT conditions for the portfolio optimization problem of Section 1.2.1.

16. Write down the KKT conditions for the optimization problem

$$\min\{f(x_1, x_2, \dots, x_n) : x_j \geq 0 \ (j \leq n), \ \sum_{j=1}^{n} x_j \leq 1\}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function.
Solution: We define $g_j(\boldsymbol{x}) = -x_j$ for $j = 1, \dots, n$, and $g_{n+1}(\boldsymbol{x}) = \sum_{j=1}^{n} x_j - 1$. We have that $\nabla g_j = -\boldsymbol{e}_j$ for $1 \leq j \leq n$, and $\nabla g_{n+1} = (1, 1, \dots, 1)$. If there are no active inequalities, we must have that $\nabla f(\boldsymbol{x}) = \mathbf{0}$. If the last constraint is not active we have that

$$\nabla f = \sum_{j \in A(x), j \leq n} \mu_j \boldsymbol{e}_j,$$

i.e. ∇f points into the cone spanned by $\boldsymbol{e}_j$, $j \in A(x)$. If the last constraint is active also , we see that

$$\nabla f = \sum_{j \notin A(x), j \leq n} -\mu_{n+1} \boldsymbol{e}_j \sum_{j \in A(x), j \leq n} (\mu_j - \mu_{n+1}) \boldsymbol{e}_j.$$

∇f is on this form whenever components outside the active set are equal and ≤ 0, and all are components are greater than or equal to this.

17. Consider the following optimization problem

$$\min\{\left(x_1 - \frac{3}{2}\right)^2 + x_2^2 : x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1\}.$$

a. Draw the region which we minimize over, and find the minimum of $f(\boldsymbol{x}) = \left(x_1 - \frac{3}{2}\right)^2 + x_2^2$ by a direct geometric argument.

b. Write down the KKT conditions for this problem. From a., decide which two conditions g_1 and g_2 are active at the minimum, and verify that you can find $\mu_1 \geq 0, \mu_2 \geq 0$ so that $\nabla f + \mu_1 \nabla g_1 + \mu_2 \nabla g_2 = \mathbf{0}$ (as the KKT conditions guarantee in a minimum) (it is not the meaning here that you should go through all possibilities for active inequalities, only those you see must be fulfilled from a.).

18. Consider the following optimization problem

$$\min\{-x_1 x_2 : x_1^2 + x_2^2 \leq 1\}$$

Write down the KKT conditions for this problem, and find the minimum.

Chapter 6

1. Consider problem (6.1) in Section 6.1. Verify that the KKT conditions for this problem are as stated there.
Solution: The constraint $A\boldsymbol{x} = \boldsymbol{b}$ actually yields one constraint per row in A, and the gradient of the i'th constraint is the i'th row in A. This gives the following sum in the KKT conditions:

$$\sum_{i=1}^{m} \nabla g_i \lambda_i = \sum_{i=1}^{m} a_{i\cdot}^T \lambda_i == \sum_{i=1}^{m} (A^T)_{\cdot i} \lambda_i = A^T \lambda.$$

The gradient of $f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{h} + \frac{1}{2}\boldsymbol{h}^T \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{h}$ is $\nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{h}$. The KKT conditions are thus $\nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{h} + A^T\lambda = \mathbf{0}$ and $A\boldsymbol{h} = \mathbf{0}$. This can be written as the set of equations

$$\begin{aligned} \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{h} + A^T\lambda &= -\nabla f(\boldsymbol{x}_k) \\ A\boldsymbol{h} + 0\lambda &= \mathbf{0}, \end{aligned}$$

from which the stated equation system follows.

2. Define the function $f(x, y) = x + y$. We will attempt to minimize f under the constraints $y - x = 1$, and $x, y \geq 0$

a. Find A, $\boldsymbol{b}$, and functions g_1, g_2 so that the problem takes the same form as in Equation (6.4).
Solution: We can set $g_1(x, y) = -x$, $g_2(x, y) = -y$, $A = \begin{pmatrix} -1 & 1 \end{pmatrix}$, and $\boldsymbol{b} = \begin{pmatrix} 1 \end{pmatrix}$.

b. Draw the contours of the barrier function $f(x, y) + \mu\phi(x, y)$ for $\mu = 0.1, 0.2, 0.5, 1$, where $\phi(x, y) = -\ln(-g_1(x, y)) - \ln(-g_2(x, y))$.

c. Solve the barrier problem analytically using the Lagrange method.
Solution: The barrier problem here is to minimize $x + y - \mu \ln x - \mu \ln y$ subject to the constraint $y - x = 1$. The gradient of the Lagrangian is $(1 - \mu/x, 1 - \mu/y) + A^T\lambda$. If this is $\mathbf{0}$ we must have that $1 - \mu/x = \mu/y - 1$, so that $2xy = \mu(x + y)$ The constraint gives that $y = x + 1$, so that $2x(x + 1) = \mu(2x + 1)$. This can be written as $2x^2 + 2(1 - \mu)x - \mu = 0$, which has the solution

$$x = \frac{-2(1-\mu) \pm \sqrt{4(1-\mu)^2 + 8\mu}}{4} = \frac{-2(1-\mu) \pm \sqrt{4 + 4\mu^2}}{4} = \frac{\mu - 1 \pm \sqrt{1 + \mu^2}}{2}.$$

If we here choose the minus sign it is straighforward to see that we get something negative, and this violates the constraint $x, y > 0$. Therefore, the barrier method obtains the minimum where $x = \frac{\mu-1+\sqrt{1+\mu^2}}{2}$, $y = x+1 = \frac{\mu+1+\sqrt{1+\mu^2}}{2}$.

d. It is straightforward to find the minimum of f under the mentioned constraints. State a simple argument for finding this minimum.
Solution: By inserting $y = x+1$ for the constraint we see that we need to minimize $g(x) = 2x+1$ subject to $x \geq 0$, which clearly has a minimum for $x = 0$, and then $y = 1$. This gives the same minimum as in c.

e. State the KKT conditions for finding the minimum, and solve these.
Solution: The KKT conditions takes one of the following forms:

- If there are no active inequalities:

$$\nabla f + A^T \lambda = (1,1) + \lambda(-1,1) = \mathbf{0},$$

  which has no solutions.

- The first inequality is active (i.e. $x = 0$):

$$\nabla f + A^T \lambda + \mu_1 \nabla g_1 = (1,1) + \lambda(-1,1) + \mu_1(-1,0) = (1-\lambda-\mu_1, 1+\lambda) = \mathbf{0},$$

  which gives that $\lambda = -1$ and $\mu_1 = \mu_1 = 2$. When $x = 0$ the equaility constraint gives that $y = 1$, so that $(0,1)$ satisfies the KKT conditions.

- The second inequality is active (i.e $y = 0$): The first constraint then gives that $x = -1$, which does not give a feasible point.

In conclusion, $(0,1)$ is the only point which satisfies the KKT conditions. If we attempt the second order test, we will see that it is inconclusive, since the Hessian of the Lagrangian is zero. To prove that $(1,0)$ must be a minimum, you can argue that f is very large outside any rectangle, so that it must have a minimum on this rectangle (the rectangle is a closed and bounded set).

f. Show that the central path converges to the same solution which you found in d. and e..
Solution: With the barrier method we obtained the solution $\boldsymbol{x}(\mu) = \left(\frac{\mu-1+\sqrt{1+\mu^2}}{2}, \frac{\mu+1+\sqrt{1+\mu^2}}{2}\right)$. Since this converges to $(0,1)$ as $\mu \to 0$, the central path converges to the solution we have found.

3. Use the function `IPBopt` to verify the solution you found in Exercise 2. Initially you must compute a feasible starting point $\boldsymbol{x}_0$.
Solution: You can use the following code:

```
IPBopt(@(x)(x(1)+x(2)),@(x)(-x(1)),@(x)(-x(2)),...
       @(x)([1;1]),@(x)([-1;0]),@(x)([0;-1]),...
       @(x)(zeros(2)),@(x)(zeros(2)),@(x)(zeros(2)),...
       [-1 1],1,[4;5])
```

4. State the KKT conditions for finding the minimum for the contstrained problem of Example 6.3, and solve these. Verify that you get the same solution as in Example 6.3.
Solution: Here we have that $\nabla f = 2x$, $\nabla g_1 = -1$, $\nabla g_2 = 1$. If there are no active constraints the KKT conditions say that $2x = 0$, so that $x = 0$, which is outside the domain of definition for f.
If the first constraint is active we get that $2x - \mu_1 = 4 - \mu_1 = 0$, so that $\mu_1 = 4$. This is a candidate for the minimum (clearly the second order conditions for a minimum is fulfilled here as well, since the Hessian of the Lagrangian is 2).
If the second constraint is active we get that $2x + \mu_2 = 4 + \mu_2 = 0$, so that $\mu_2 = -4$, so that this gives no candidate for a solution.
It is impossible for both constraints to be active at the same time, so $x = 2$ is the unique minimum.

5. In the function `IPBopt2`, replace the call to the function `newtonbacktrackg1g2` with a call to the function `newtonbacktrack`, with the obvious modification to the parameters. Verify that the code does not return the expected minimum in this case.

6. Consider the function $f(x) = (x-3)^2$, with the same constraints $2 \le x \le 4$ as in Example 6.3. Verify in this case that the function `IPBopt2` returns the correct minimum regardless of whether you call `newtonbacktrackg1g2` or `newtonbacktrack`. This shows that, at least in some cases where the minimum is an interior point, the iterates from Newtons method satisfy the inequality constraints as well.
Solution: You can use the following code:

```
IPBopt2(@(x)((x-3).^2),@(x)(2-x),@(x)(x-4),...
       @(x)(2*(x-3)),@(x)(-1),@(x)(1),...
       @(x)(2),@(x)(0),@(x)(0),3.5)
```

7. (Trial Exam UIO V2012) In this exercise we will find the minimum of the function $f(x, y) = 3x + 2y$ under the constraints $x + y = 1$ and $x, y \ge 0$.

a. Find a matrix A and a vector $\boldsymbol{b}$ so that the constraint $x + y = 1$ can be written on the form $A\boldsymbol{x} = \boldsymbol{b}$.
Solution: We can set $A = \begin{pmatrix} 1 & 1 \end{pmatrix}$, og $\boldsymbol{b} = 1$.

b. State the KKT-conditions for this problem, and find the minimum by solving these.
Solution: We set $g_1(x, y) = -x \le 0$ and $g_2(x, y) = -y \le 0$, and have that

$\nabla f = (3,2)$, $\nabla g_1 = (-1,0)$, $\nabla g_2 = (0,-1)$. The KKT-conditions therefore take the form $x + y = 1$ and

$$\nabla f + A^T\lambda + \nu_1\nabla g_1 + \nu_2\nabla g_2 = (3,2) + \lambda(1,1) + \nu_1(-1,0) + \nu_2(0,-1) = \mathbf{0},$$

where the two last terms are included only if the corresponding inequalities are active, and where $\nu_1, \nu_2 \geq 0$.
If none of the inequalities are active we get that $(3,2) + \lambda(1,1) = \mathbf{0}$, which has now solution.
If both inequalities are active we get that $x = y = 0$, which does not fulfill the constraint $x + y = 1$. If we have only one active inequality we have two possibilities: If the first inequality is active we get that $(3,2) + \lambda(1,1) + \nu_1(-1,0) = \mathbf{0}$. The equation for the second component says that $\lambda = -2$, and the equation for the first component says that $3 - 2 - \nu_1 = 0$, so that $\nu_1 = 1$.
If the second inequality is active we get that $(3,2) + \lambda(1,1) + \nu_2(0,-1) = \mathbf{0}$. The equation for the first component says that $\lambda = -3$, and the equation for the second component says that $2 - 3 - \nu_2 = 0$, which gives that $\nu_2 = -1$. This possibilitywe must denounce since $\nu_2 < 0$. We are left with the first inquality as active as the only possibility. Then $x = 0$, and the constraint $x + y = 1$ gives that $y = 1$, and a minimum value of 2. Since f clearly is bounded below on the region we work on, it is clear that this must be a global minimum.

Finally we should candidates for the minimum which are not regular points. If none of the equations are active we have no candidates, since $\nabla h_1 = (1,1) \neq \mathbf{0}$. If one inequality is active we get no candidates either, since $(1,1)$ and $(-1,0)$ are linearly independent, and since $(1,1)$ and $(0,-1)$ are linearly independent. If both inequalities are active ($x = y = 0$), it is clear that the constraint $x+y = 1$ is not fulfilled. All in all, we get no additional candidates from points which are not regular.

c. Write down the barrier function $\phi(x,y) = -\ln(-g_1(x,y)) - \ln(-g_2(x,y))$ for this problem, where g_1 and g_2 represent the two constraints of the problem. Also compute $\nabla\phi$.
Solution: We get that $\phi(x,y) = -\ln x - \ln y$, and $\nabla\phi = (-1/x, -1/y)$.

d. Solve the barrier problem with parameter μ, and denote the solution by $\boldsymbol{x}(\mu)$. Is it the case that the limit $\lim_{\mu\to 0}\boldsymbol{x}(\mu)$ equals the solution you found in b.?
Solution: In the barrier problem we minimize the function $f(x,y)+\mu\phi(x,y) = 3x+2y-\mu\ln x-\mu\ln y$ under the constraint $x + y = 1$. The KKT-conditions become $(3,2) + \mu(-1/x, -1/y) + \lambda(1,1) = \mathbf{0}$, which gives the equations

$$\begin{aligned}\frac{\mu}{x} &= 3 + \lambda \\ \frac{\mu}{y} &= 2 + \lambda.\end{aligned}$$

This gives that $\frac{\mu}{y} + 1 = \frac{\mu}{x}$, which again gives $\mu(y - x) = xy$. If we substitute the constraint $x + y = 1$ we get that $\mu(1-2x) = x(1-x)$, which can be written

$x^2 - (1+2\mu)x + \mu = 0$. If we solve this we find that

$$x = \frac{1+2\mu \pm \sqrt{(1+2\mu)^2 - 4\mu}}{2} = \frac{1+2\mu \pm \sqrt{1+4\mu^2}}{2}.$$

This corresponds to two different points, depending on which sign we choose, but if we choose + as sign we see that $x > 1$, so that $y < 0$ in order for $x + y = 1$, so that (x, y) then is outside the domain of definition for the problem. We therefore have that $x = \frac{1+2\mu-\sqrt{1+4\mu^2}}{2}$. It is clear that $x \to 0$ when $\mu \to 0$ here, so that the solution of the barrier problem converges to the solution of the original problem.