Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimisation I Spring 2018

Solutions to exercise set 10

1 Consider the constrained optimisation problem

$$\frac{1}{2}(x^2+y^2) \to \min \qquad \text{subject to } xy = 1.$$

a) Find—by whatever means—the solutions of this problem. In addition, find the values of the corresponding Lagrange multipliers.

Solution: One strategy: let $f(x,y) = \frac{1}{2}(x^2+y^2)$ and $c(x,y) = xy-1$. Clearly

$$f(x,y) = \tfrac{1}{2}(x-y)^2 + xy = \tfrac{1}{2}(x-y)^2 + 1,$$

whose global minimisers evidently satisfy $x = y$. And from the constraint $xy = 1$, this gives solutions $(-1,-1)$ and $(1,1)$. Furthermore, at optima, ∇f must be parallel to ∇c, or, $\nabla f = \lambda \nabla c$ for some Lagrange multiplier $\lambda \in \mathbb{R}$. Since $\nabla f(-1,-1) = (-1,-1)$ and $\nabla c(-1,-1) = (-1,-1)$, this gives $\lambda = 1$. At $(1,1)$, we similarly find a corresponding $\lambda = 1$.

(Another option is to set up and solve the KKT conditions, plus argue, for example, via second order sufficient conditions that these points are indeed minima.)

b) Formulate the unconstrained optimisation problem that results from the application of the quadratic penalty method with parameter $\mu > 0$. Solve these problems for all possible parameters μ and verify that the solutions converge to the solutions of the constrained optimization problem as $\mu \to \infty$.

Solution: The quadratic penalty method with parameter $\mu > 0$ seeks to minimise

$$Q(x,y;\mu) := f(x,y) + \frac{\mu}{2}c(x,y)^2 = \frac{1}{2}\big(x^2+y^2\big) + \frac{\mu}{2}(xy-1)^2$$

unconstrained over all $(x,y) \in \mathbb{R}^2$. Note that Q is smooth and coercive and thus admits a global minimum, which also must be a stationary point. Calculating

$$\nabla Q(x,y;\mu) = \begin{bmatrix} x + \mu(xy-1)y \\ y + \mu(xy-1)x \end{bmatrix},$$

we find that the first component of ∇Q vanishes whenever

$$x = \frac{\mu y}{1+\mu y^2}.$$

Inserted into the second component of the equation $\nabla Q = 0$, this yields

$$y\left[1 - \frac{\mu^2}{(1+\mu y^2)^2}\right] = 0. \tag{$\star$}$$

If $y = 0$, then $x = 0$ also, so $(0,0)$ is a stationary point. Examining the Hessian of Q at $(0,0)$ shows that $\nabla^2 Q(0,0;\mu)$ is positive definite when $\mu < 1$, and negative definite when $\mu > 1$. Thus $(0,0)$ is a strict local minimiser when $\mu < 1$ and a strict local maximiser when $\mu > 1$. If $\mu = 1$, then

$$Q(x,y;1) = \frac{1}{2}\left[(x-y)^2 + (xy)^2 + 1\right] \geq \frac{1}{2} = Q(0,0),$$

with equality if and only if $x = y = 0$. As such, $(0,0)$ is a strict local minimiser also for $\mu = 1$.
If $y \neq 0$, then ($\star$) simplifies to

$$1 + \mu y^2 = \mu,$$

with solutions

$$y = \pm\sqrt{1 - \frac{1}{\mu}},$$

provided $\mu \geq 1$. This also gives

$$x = \frac{\mu y}{1+\mu y^2} = \pm\sqrt{1-\frac{1}{\mu}},$$

and it can be verified (how?) that these points (x,y) are minimisers.
In total, $(0,0)$ is the global minimiser of $Q(\cdot,\cdot;\mu)$ when $\mu \leq 1$, while the two points

$$(x,y) = \left(\pm\sqrt{1-\frac{1}{\mu}}, \pm\sqrt{1-\frac{1}{\mu}}\right)$$

minimise $Q(\cdot,\cdot;\mu)$ when $\mu > 1$. Finally, as $\mu \to \infty$, we find that (x,y) converges to the global minimisers $(\pm 1, \pm 1)$ of the original constrained problem.

c) Formulate the augmented Lagrangian for this constrained optimization problem and find (for all possible parameters $\lambda \in \mathbb{R}$ and $\mu > 0$) the global solutions of this (unconstrained) optimization problem. For which parameters does one recover the solution of the original constrained problem?

Solution: The augmented Lagrangian for this problem is

$$L_A(x,y,\lambda,\mu) = \frac{1}{2}(x^2+y^2) - \lambda(xy-1) + \frac{\mu}{2}(xy-1)^2,$$

which is coercive and lower semi-continuous such that a minimizer exists, and it has gradient

$$\nabla L_A(x,y,\lambda,\mu) = \begin{bmatrix} x - \lambda y + \mu(xy^2 - y) \\ y - \lambda x + \mu(x^2 y - x) \end{bmatrix}.$$

After a similar computation to that in part b), we find

$$x = \frac{(\mu+\lambda)y}{1+\mu y^2}$$

and the equation for y:

$$(1 + \mu y^2)^2 = (\lambda + \mu)^2.$$

In addition, we have the solution $(x, y) = (0, 0)$. We must be somewhat careful in finding y. First, we have

$$1 + \mu y^2 = \pm(\lambda + \mu),$$

but since the left hand side is positive, we must choose the right hand side positive as well. Therefore, we have

$$1 + \mu y^2 = |\lambda + \mu|$$

and thus

$$y^* = \pm\sqrt{\left|\frac{\lambda}{\mu} + 1\right| - \frac{1}{\mu}},$$

which exist if $|\lambda + \mu| \geq 1$. It can be checked that here, too, we have $x^* = y^*$. The points (x^*, y^*) are the global minimizers if $\lambda + \mu \geq 1$. Otherwise, $(0, 0)$ is the global minimizer. We see that the original solution is obtained when either $\lambda = 1$ or $\mu \to \infty$. The fact that (x^*, y^*) are the global minimizers if $\lambda + \mu \geq 1$ can seen by checking when $\mathcal{L}_A(x^*, y^*, \lambda, \mu) \leq \mathcal{L}_A(0, 0, \lambda, \mu)$. This leads (after some computation) to the condition

$$(\lambda + \mu - 1)(|\lambda + \mu| - 1) \geq \frac{1}{2}(|\lambda + \mu| - 1)^2.$$

Since (x^*, y^*) exist only if $|\lambda + \mu| \geq 1$, and if $|\lambda + \mu| = 1$ then $(x^*, y^*) = (0, 0)$, we can divide by $|\lambda + \mu| - 1$ to obtain the condition

$$(\lambda + \mu - 1) \geq \frac{1}{2}(|\lambda + \mu| - 1),$$

which holds if $\lambda + \mu \geq 1$ but not if $\lambda + \mu \leq -1$.

d) The ℓ^1-penalty function for this optimisation problem is defined, for some parameter $\mu > 1$, as

$$\Phi_1(x, y; \mu) := \frac{1}{2}(x^2 + y^2) + \mu|xy - 1|.$$

Find for each parameter $\mu > 0$ the global minimisers of this function. For which parameters $\mu > 0$ do they coincide with the solutions of the original problem?

Solution: We find the minimizers of

$$\Phi_1(x, y; \mu) = \frac{1}{2}(x^2 + y^2) + \mu|xy - 1|$$

by splitting the domain in three: $xy > 1$, $xy < 1$ and $xy = 1$. First, when $xy = 1$, we see that $x^2 = 1/y^2$, so the objective function takes the form

$$\Phi_1(x, y; \mu) = g(y) = \frac{1}{2}\left(\frac{1}{y^2} + y^2\right).$$

We can see that $g'(y) = 0$ when $y = 1$ or $y = -1$, and $g''(y) = 4$ in both these points, so they are minimizers along the curves $x = 1/y$, and we have the candidates (-1,-1) and (1,1). Furthermore, $\Phi_1(1, 1; \mu) = \Phi_1(-1, -1; \mu) = 1$ for all values of μ.
Next, if $xy > 1$ then

$$\Phi_1(x, y; \mu) = \frac{1}{2}(x^2 + y^2) + \mu(xy - 1),$$

so

$$\nabla\Phi_1(x, y; \mu) = \begin{bmatrix} x + \mu y \\ y + \mu x \end{bmatrix} = 0 \Rightarrow x = -\mu y \Rightarrow (1 - \mu^2)y = 0.$$

If $y = 0$ then $x = 0$, but this is not in the domain considered so we need to take $\mu = \pm 1$. Since $\mu > 0$, the only possibility is $\mu = 1$. This gives us the critical points along the line $x = -y$, but this is still not in the domain considered. Thus, there are no critical points in the domain $xy > 1$.

Finally, in the domain $xy < 1$, we have

$$\Phi_1(x, y; \mu) = \frac{1}{2}(x^2 + y^2) - \mu(xy - 1),$$

so

$$\nabla\Phi_1(x, y; \mu) = \begin{bmatrix} x - \mu y \\ y - \mu x \end{bmatrix} = 0 \Rightarrow x = \mu y \Rightarrow (1 - \mu^2)y = 0.$$

If $y = 0$ then $x = 0$. This is in the domain and thus a critical point. Also, we may take $\mu = \pm 1$. Since $\mu > 0$, the only possibility is $\mu = 1$. This gives us the critical points along the line $x = y$, which are in the domain considered when $|x| < 1$. We now check whether any of these points are minimizers. Observe that

$$\nabla^2\Phi_1(x, y; \mu) = \begin{bmatrix} 1 & -\mu \\ -\mu & 1 \end{bmatrix}$$

with eigenvalues $\lambda = 1 \pm \mu$. The eigenvalues are positive when $\mu < 1$ and so the point (0,0) is a local minimizer when $\mu < 1$, with value $\Phi_1(0, 0; \mu) = \mu$, which actually makes it a global minimizer.

When $\mu = 1$, we have $\Phi_1(x, y; \mu) = 1$ along the line $x = y$. Also, when $\mu = 1$, we have

$$\Phi_1(x, y; \mu) = \frac{1}{2}(x - y)^2 + 1 \geq 1,$$

so these points are minimizers.

When $\mu > 1$, the global minimizers are found in $(x, y) = (\pm 1, \pm 1)$. This is because $\Phi_1(\pm 1, \pm 1, \mu) = 1$ and $\Phi_1(x, y; \mu) > 1$ elsewhere. This can be seen as following: When $xy > 1$,

$$\begin{aligned} \Phi_1(x, y; \mu) &= \frac{1}{2}(x^2 + y^2) + \mu(xy - 1) \\ &= \frac{1}{2}(x - y)^2 + \mu(xy - 1) + xy \\ &\geq \mu(xy - 1) + xy \\ &> 1, \end{aligned}$$

and when $xy < 1$:

$$\begin{aligned}\Phi_1(x, y; \mu) &= \frac{1}{2}(x^2 + y^2) - \mu(xy - 1) \\ &= \frac{1}{2}(x - y)^2 + xy - \mu(xy - 1) \\ &\geq \mu(1 - xy) + xy \\ &> 1.\end{aligned}$$

To summarize: When $\mu < 1$, we have a global minimizer in $(0, 0)$ with value μ. When $\mu = 1$, the global minimizers can be found on the line $x = y, x \in [-1, 1]$, and with $\mu > 1$, the global minimizers are found in $(x, y) = (\pm 1, \pm 1)$.

2 Consider the constrained optimisation problem

$$x + y \to \min \qquad \text{subject to } x^2 + y^2 \leq 1.$$

Formulate a logarithmic barrier method for the solution of this constrained optimisation problem and compute its solution for each parameter $\mu > 0$ in the barrier functional.

Solution: The logarithmic barrier approach may be written as

$$\min_{x,y,s} x + y - \mu \log s \qquad \text{subject to} \qquad 1 - x^2 - y^2 - s = 0,$$

where s (≥ 0) is the slack variable, and $\mu > 0$ is the barrier parameter which we intend to drive to 0. Introducing a Lagrange multipler λ, the KKT conditions for this problem are

$$1 + 2x\lambda = 0, \qquad 1 + 2y\lambda = 0, \qquad -\frac{\mu}{s} + \lambda = 0, \qquad \text{and} \qquad 1 - x^2 - y^2 - s = 0.$$

This gives first that

$$\lambda = \frac{\mu}{s} \qquad \text{and} \qquad x = y = -\frac{s}{2\mu},$$

and inserted into the constraint equation, we find that

$$1 - \frac{s^2}{2\mu^2} - s = 0.$$

The relevant solution of this quadratic equation is $s = \mu\big(\sqrt{\mu^2 + 2} - \mu\big)$, and we end up with

$$x = y = -\frac{1}{2}\big(\sqrt{\mu^2 + 2} - \mu\big) \qquad \text{and} \qquad \lambda = \big(\sqrt{\mu^2 + 2} - \mu\big)^{-1}.$$

Since the Hessian of the Lagrangian to this problem is positive definite (check it!), the found KKT point is the unique global minimiser of the logarithmic barrier formulation. Notably, as $\mu \to 0^+$, we recover the exact solution $x^* = y^* = -1/\sqrt{2}$, with $\lambda^* = 1/\sqrt{2}$, of the original problem.

3 Consider the quadratic programming problem

$$\min (x+1)^2 + (y+1)^2, \qquad \text{subject to} \quad \begin{aligned} x+y &\geq 0,\\ -x &\geq -2,\\ -y &\geq -2. \end{aligned}$$

a) Find the global minimum for this optimization problem (you can do this graphically). Determine the corresponding active set and Lagrange multipliers.

Solution: The objective function is the distance between feasible points and the point $(-1,-1)$. Clearly the minimum is attained at $(x^*,y^*) = (0,0)$. At this point the active set is $\mathcal{A}^* = \{1\}$, and therefore from complementarity we know that $\lambda_2^* = \lambda_3^* = 0$. λ_1 is determined from the equation $\nabla f(x^*,y^*) = \lambda_1^* \nabla c_1(x^*,y^*)$, from which $\lambda_1^* = 2$.

b) Solve the problem using an active set method. Start with $(x_0,y_0) = (2,0)$ and working set $W_0 = \{2\}$.

Solution: We only outline the computations.

At the first iteration, the minimum of f restricted to the constraint $x = 2$ occurrs at $(\tilde{x}_1,\tilde{y}_1) = (2,-1)$; equivalently this determines the first search direction $(p_1,q_1) = (2,-1)-(2,0) = (0,-1)$. Since the point $(\tilde{x}_1,\tilde{y}_1)$ is feasible for our problem we accept the unit step and put $(x_1,y_1) = (2,-1)$ with $W_1 = W_0$.

With the current working set we cannot progress any longer (we are already at the point of minimum along this constraint), so we look at the Lagrange multiplier(s), which in this case is -6. Since it is negative, we drop 2 from the set of active constraints and solve the unconstrained problem (empty working set) at the next step.

The unconstrained problem has a solution $(-1,-1)$, which is infeasible. Thus we cannot take a unit step along the direction $(-1,-1)-(2,-1) = (-3,0)$. The binding constraint (this is cleat graphically) is $x+y \geq 0$, thus the new point is going to be when this constraint becomes active along our search direction. This happens at the point $(x,y) = (1,-1)$ correspoinding to the steplength $\alpha = 1/3$. Thus our new working set is $W = \{1\}$.

At this stage we have found the optimal set (the algorithm just does not know it yet). Minimizing the objective along the line $x+y = 0$ produces the global optimum $(0,0)$ (corresponds to the search direction $(-1,1)$ and the unit step along it.

4 Assume that $A \in \mathbb{R}^{m\times n}$ with $m < n$ is a matrix of full rank and that $b \in \mathbb{R}^m \setminus \{0\}$. Consider the optimisation problem

$$\frac{1}{2}\|x\|^2 \to \min \qquad \text{subject to} \qquad Ax = b. \tag{1}$$

(See also problem 3 in exercise set 7.)

a) Formulate the augmented Lagrangian $\mathcal{L}_A$ for problem (1), and find for all possible parameters $\lambda \in \mathbb{R}^m$ and $\mu > 0$ a formula for the global solutions of the resulting unconstrained optimisation problem.

Solution: In matrix form—since we have m constraints—the augmented Lagrangian equals

$$\mathcal{L}_A(x, \lambda; \mu) = \frac{1}{2}\|x\|^2 - \lambda^\top (Ax - b) + \frac{\mu}{2}\|Ax - b\|^2,$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multipler, and $\mu > 0$ is the penalty parameter. Trying to minimise $\mathcal{L}_A$ with respect to x, we first note that $\mathcal{L}_A(\cdot, \lambda; \mu)$ is both smooth and coercive, where the latter property follows from the dominating $\frac{1}{2}\|x\|^2$ term (how?). Hence, a global minimiser exists. This point is also a stationary point, and computing

$$0 = \nabla_x \mathcal{L}_A(x, \lambda; \mu) = x - A^\top \lambda + \mu A^\top (Ax - b)$$

shows that the unique global minimiser, for all λ and all μ, is

$$\begin{aligned} x_{\lambda,\mu} &= \Big(\frac{1}{\mu}\mathrm{Id} + A^\top A\Big)^{-1} A^\top \Big(\frac{1}{\mu}\lambda + b\Big) \\ &= A^\top \Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \Big(\frac{1}{\mu}\lambda + b\Big). \end{aligned}$$

(Last transition was shown in exercise 3 b) in exercise set 7.)

b) Since the exact solution and optimal Lagrange multiplier of the original optimisation problem equal

$$x^* = (A^\top A)^{-1} A^\top b = A^\top (AA^\top)^{-1} b = A^\top \lambda^* \qquad \text{and} \qquad \lambda^* = (AA^\top)^{-1} b,$$

we demand that $x_{\lambda,\mu} = x^*$ and see what happens: left-multipling both sides by $(AA^\top)^{-1}A$ gives

$$\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \Big(\frac{1}{\mu}\lambda + b\Big) = \lambda^*,$$

or,

$$\frac{1}{\mu}\lambda + b = \Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)\lambda^* = \frac{1}{\mu}\lambda^* + b.$$

In conclusion, the minimiser of the augmented Lagrangian equals that of the exact solution if and only if $\lambda = \lambda^*$, with no restrictions on $\mu > 0$.

c) For which parameters $\lambda \in \mathbb{R}^m$ and $\mu > 0$ is the minimiser of the augmented Lagrangian equal to the solution of (1)?

Solution: Before we should make any use of the iterative algorithm, it is vital to establish its *consistency* with the original optimisation problem which it intends to solve. By this we mean that the algorithm should solve the original problem in the limit: if $x^k \to x$ and $\lambda^k \to \lambda$, then $x = x^*$ and $\lambda = \lambda^*$.
As x^{k+1} is the minimum of $\mathcal{L}_A(\cdot, \lambda^k; \mu)$, we know that $x^{k+1} = x_{\lambda^k,\mu}$, and so inserting this into the iteration for the Lagrange multiplier yields that

$$\lambda^{k+1} = \lambda^k - \mu\Big(Ax^{k+1} - b\Big) = M(\lambda^k + \mu b),$$

where

$$\begin{aligned}M &= \mathrm{Id} - AA^\top \Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \Big[\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big) - AA^\top\Big]\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \frac{1}{\mu}\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \big(\mathrm{Id} + \mu AA^\top\big)^{-1}.\end{aligned}$$

Suppose now that $x^k \to x$ and $\lambda^k \to \lambda$. Then from the Lagrange multiplier iteration we get

$$\lambda = M(\lambda + \mu b),$$

which implies that

$$\big(\mathrm{Id} + \mu AA^\top\big)\lambda = \lambda + \mu b.$$

In other words,

$$\lambda = (AA^\top)^{-1} b = \lambda^*,$$

and therefore $x = x^*$ as well, because x^* is the minimiser of $\mathcal{L}_A(\cdot, \lambda^*; \mu)$ from question b). Thus the scheme is consistent.

Recall now from numerical linear algebra that the consistent Lagrange multiplier iteration will converge for all initial values if and only if $\rho(M) < 1$, where $\rho(M)$ denotes the spectral radius of M, that is, the largest eigenvalue of M in absolute value. Since A has full rank, matrix $AA^\top$ is positive definite, with strictly positive eigenvalues. As such, the eigenvalues of

$$M^{-1} = \mathrm{Id} + \mu AA^\top$$

are always strictly greater than 1, which means that $\rho(M) < 1$, as desired.

d) An iterative algorithm for the solution of (1) using the augmented Lagrangian may have the form

$$\begin{aligned}x^{k+1} &\in \arg\min_x \mathcal{L}_A(x, \lambda^k; \mu), \\ \lambda^{k+1} &= \lambda^k - \mu(Ax^{k+1} - b).\end{aligned}$$

Show that this iteration converges for all initial values $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$, and all $\mu > 0$ to the unique solution of (1).[1]

Hint: Interpret the iteration for the Lagrange parameter λ as a fixed-point iteration—by using the explicit formula for x^{k+1} derived in the first part of this exercise—and then use results from previous numerics courses to show that this fixed-point scheme converges.

Solution: Before we should make any use of the iterative algorithm, it is vital to establish its *consistency* with the original optimisation problem which it intends to solve. By this we mean that the algorithm should solve the original problem in the limit: if $x^k \to x$ and $\lambda^k \to \lambda$, then $x = x^*$ and $\lambda = \lambda^*$.

[1] This problem does not completely fall within the curriculum of this optimisation class, but it is still recommended to *try* to solve it.

As x^{k+1} is the minimum of $\mathcal{L}_A(\cdot, \lambda^k; \mu)$, we know that $x^{k+1} = x_{\lambda^k,\mu}$, and so inserting this into the iteration for the Lagrange multiplier yields that

$$\lambda^{k+1} = \lambda^k - \mu\left(Ax^{k+1} - b\right) = M(\lambda^k + \mu b),$$

where

$$\begin{aligned} M &= \mathrm{Id} - AA^\top \Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \left[\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big) - AA^\top\right]\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \frac{1}{\mu}\Big(\frac{1}{\mu}\mathrm{Id} + AA^\top\Big)^{-1} \\ &= \big(\mathrm{Id} + \mu AA^\top\big)^{-1}. \end{aligned}$$

Suppose now that $x^k \to x$ and $\lambda^k \to \lambda$. Then from the Lagrange multiplier iteration we get

$$\lambda = M(\lambda + \mu b),$$

which implies that

$$\big(\mathrm{Id} + \mu AA^\top\big)\lambda = \lambda + \mu b.$$

In other words,

$$\lambda = (AA^\top)^{-1}b = \lambda^*,$$

and therefore $x = x^*$ as well, because x^* is the minimiser of $\mathcal{L}_A(\cdot, \lambda^*; \mu)$ from question b). Thus the scheme is consistent.

Recall now from numerical linear algebra that the consistent Lagrange multiplier iteration will converge for all initial values if and only if $\rho(M) < 1$, where $\rho(M)$ denotes the spectral radius of M, that is, the largest eigenvalue of M in absolute value. Since A has full rank, matrix $AA^\top$ is positive definite, with strictly positive eigenvalues. As such, the eigenvalues of

$$M^{-1} = \mathrm{Id} + \mu AA^\top$$

are always strictly greater than 1, which means that $\rho(M) < 1$, as desired.

Useful matrix formula: if B, C, and $B + C$ are invertible matrices of the same size, then

$$\big(B^{-1} + C^{-1}\big)^{-1} = B(B + C)^{-1}C = C(B + C)^{-1}B.$$