Norwegian University of Science
and Technology
Department of Mathematical
Sciences

TMA4180
Optimization I
Spring 2021

Solutions to exercise set 7

1 Consider the constrained optimization problem

$$xy \to \min \qquad \text{such that} \qquad \begin{cases} y \geq x, \\ y^4 \leq x^3. \end{cases}$$

a) Find all KKT points and local minima for this optimization problem.

Solution: We begin, as usual, by stating the problem in standard form, writing $\mathbf{x} = [x, y]^T$:

$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \quad \text{s.t.} \quad c_i(\mathbf{x}) \geq 0, \quad i = 1, 2$$

where

$$\begin{aligned} f(\mathbf{x}) &= xy, \\ c_1(\mathbf{x}) &= y - x \\ c_2(\mathbf{x}) &= x^3 - y^4 \end{aligned}$$

The KKT conditions are:

$$y + \lambda_1 - 3x^2\lambda_2 = 0 \tag{1a}$$
$$x - \lambda_1 + 4y^3\lambda_2 = 0 \tag{1b}$$
$$y - x \geq 0 \tag{1c}$$
$$x^3 - y^4 \geq 0 \tag{1d}$$
$$\lambda_i \geq 0, \quad i = 1, 2 \tag{1e}$$
$$\lambda_1(y - x) = 0 \tag{1f}$$
$$\lambda_2(x^3 - y^4) = 0. \tag{1g}$$

Now, we can check the different cases of active constraints to find KKT points. First, with no active constraints, i.e. $\lambda_1 = \lambda_2 = 0$, we get the point (0,0). In fact, this is a point with both constraints active; it just so happens that $\lambda_1 = \lambda_2 = 0$ here. One can check that the LICQ does not hold here, but it is still a KKT point because the gradient of f at (0,0) (which is 0) can be written as a non-negative linear combination of the gradients of the constraints. Thereby, we conclude that there exist no KKT points with no active constraints, but that (0,0) is a KKT point. Note that without CQ KKT conditions no longer *have to* hold at points of local minima, but they *may* still hold!

Next, we check with one active constraint. First, with $\lambda_1 \geq 0, \lambda_2 = 0$, we

have from (1a) and (1b) that $x = \lambda_1$ and $y = -\lambda_1$. Inserting into the equality (1c) yields $\lambda_1 = 0$, and therefore $(x, y) = (0, 0)$ again, which has been discussed already.

With $\lambda_2 \geq 0, \lambda_1 = 0$, equations (1a), (1b) and (1d) become

$$\begin{aligned} y - 3x^2\lambda_2 &= 0, \\ x + 4y^3\lambda_2 &= 0, \\ x^3 &= y^4. \end{aligned}$$

Multiplying the first of these by x, the second by y, applying the third and adding the two first gives

$$y^4\lambda_2 = 0.$$

Any solution of this leads to the point (0,0), which we have already found to be a KKT point.

Finally, we check the case with two active constraints, for which there are two points; (0,0), which is already considered, and (1,1). In the point (1,1), we find (by (1a) and (1b)) that $\lambda_1 = -7$ and $\lambda_2 = -2$ and is therefore this is not a KKT point.
Thus the only KKT point is $(0, 0)$. This also happens to be the only point, where CQ do not hold. Since the continuous function must attain its minimum on a non-empty bounded and closed set, and KKT conditions are necessary for optimality at all other feasible points, the point $(0, 0)$ must the the point of global, hence also local minimum. Indeed: all other feasible points in our domain have positive coordinates, and therefore the minimum of xy occurrs at $(x, y) = (0, 0)$.

b) Compute the critical cone at $(0, 0)$, and show that there exist directions d contained in the critical cone for which $d^\top \nabla^2 \mathcal{L}((0, 0); \lambda^*)d < 0$.

Solution: To find the critical cone C at $(0, 0)$, we use the definition given by equation (12.53) in N&W page 330. First, we find the gradients of the constraints at this point:

$$\nabla c_1(0, 0) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \qquad \nabla c_2(0, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $\lambda_1 = \lambda_2 = 0$ at this point, we have that $d \in C(0, 0)$ if and only if $\nabla c_1(0, 0)^T d \geq 0$ and $\nabla c_2(0, 0)^T d \geq 0$. The latter condition clearly holds for all d, and so we find that

$$\begin{aligned} C(0, 0) &= \{d = (d_1, d_2) \in \mathbb{R}^2 : \nabla c_1(0, 0)^T d \geq 0\} \\ &= \{d = (d_1, d_2) \in \mathbb{R}^2 : d_2 \geq d_1\}. \end{aligned}$$

Next, we find that the Hessian of the Lagrangian at $(0, 0)$ with Lagrange multipliers $\lambda^* = (\lambda_1, \lambda_2) = (0, 0)$ is given by

$$\nabla^2 \mathcal{L}((0, 0); (\lambda_1, \lambda_2)) = \nabla^2 \mathcal{L}((0, 0); (0, 0)) = \nabla^2 f(0, 0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Thus, $d^T \nabla^2 \mathcal{L}((0, 0); (\lambda_1, \lambda_2))d = 2d_1 d_2$. There are clearly directions in the critical cone for which this is negative; one can choose $d_2 > 0$ and $d_1 < 0$.

c) Show that $d^\top \nabla^2 \mathcal{L}((0,0);\lambda^*)d \geq 0$ for all vectors d contained in the tangent cone to the feasible set at $(0,0)$.

Solution: We can find the tangent cone to the feasible set at $(0,0)$ by looking at the limiting vectors along the lines $c_1(\mathbf{x}) = 0$ and $c_2(\mathbf{x}) = 0$ as $\mathbf{x} \to 0$. Traveling toward (0,0) along $c_1(\mathbf{x}) = 0$ we take, for example,

$$z_k = \begin{bmatrix} 1/k \\ 1/k \end{bmatrix}, \quad t_k = ||z_k|| = \sqrt{2}/k,$$

and find the limiting direction

$$d = \lim_{k\to\infty} \frac{z_k}{t_k} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Along $c_2(\mathbf{x}) = 0$, we take

$$z_k = \begin{bmatrix} 1/k \\ 1/k^{3/4} \end{bmatrix}, \quad t_k = ||z_k|| = \frac{\sqrt{\sqrt{k}+1}}{k},$$

and find the limiting direction

$$d = \lim_{k\to\infty} \frac{z_k}{t_k} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The tangent cone at (0,0) contains all vectors between these limiting cases, which can be shown to be:

$$T(0,0) = \{d \in \mathbb{R}^2 : d_1 \geq 0 \text{ and } d_2 \geq d_1\}.$$

It is then easy to see that $d^T \nabla^2 \mathcal{L}((0,0);(\lambda_1,\lambda_2))d = 2d_1 d_2 \geq 0$ for all d in the tangent cone at $(0,0)$.

2 Consider the constrained optimization problem

$$-x^2 - (y-1)^2 \to \min \qquad \text{such that} \qquad \begin{cases} y \geq Cx^2, \\ y \leq 2, \end{cases}$$

where $C > 0$ is some positive parameter.

a) Show that the point $(0,0)$ is a KKT point for all parameters $C > 0$ and that the LICQ is satisfied at $(0,0)$.

Solution: Introducing

$$f(x,y) = -x^2 - (y-1)^2, \quad c_1(x,y) = y - Cx^2, \qquad \text{and} \quad c_2(x,y) = 2 - y,$$

the minimisation problem becomes

$$\min_{x,y} f(x,y) \qquad \text{subject to} \qquad c_1(x,y) \geq 0 \text{ and } c_2(x,y) \geq 0.$$

Let also

$$\mathcal{L}(x,y,\lambda_1,\lambda_2) = f(x,y) - \lambda_1 c_1(x,y) - \lambda_2 c_2(x,y)$$

be the Lagrangian, with multipliers λ_1 and λ_2.
Focusing on the KKT conditions, it is clear that $(0,0)$ is feasible. Moreover, from the complementarity conditions

$$\lambda_1 c_1(0,0) = 0 \qquad \text{and} \qquad \lambda_2 c_2(0,0) = 0,$$

we require $\lambda_2 = 0$ because c_2 is inactive. (Note: c_1 is active, so the first condition holds.) Computing

$$\nabla_{x,y}\mathcal{L}(x,y,\lambda_1,0) = \nabla f(x,y) - \lambda_1 \nabla c_1(x,y) = \begin{bmatrix} -2x(1-\lambda_1 C) \\ -2(y-1) - \lambda_1 \end{bmatrix}$$

and demanding that this gradient vanishes at $(x,y) = (0,0)$, then give $\lambda_1 = 2$, with no restriction on C. Hence, $(0,0)$ is a KKT point for all $C > 0$. Additionally, since c_1 is the only active constraint and $\nabla c_1(0,0) = (0,1) \neq 0$, it follows that the LICQ is satisfied as well.

b) Formulate the second order necessary and sufficient optimality conditions for the point $(0,0)$. For which parameters C are these conditions satisfied? For which parameters C is the point $(0,0)$ a local minimum?

Solution: Let $\mathcal{C}$ be the critical cone at $(0,0)$ with Lagrange multipliers $(\lambda_1,\lambda_2) = (2,0)$. Then $(0,0)$, which satisfies the KKT conditions, is a local minimiser of the constrained problem only if (*necessary condition*) the Lagrangian Hessian

$$\nabla^2_{(x,y)}\mathcal{L}(0,0,2,0) = \begin{bmatrix} -2(1-2C) & 0 \\ 0 & -2 \end{bmatrix}$$

is positive semi-definite on $\mathcal{C}$, that is,

$$w^\top \nabla^2_{(x,y)}\mathcal{L}(0,0,2,0) w \geq 0 \quad \text{for all} \quad w \in \mathcal{C}.$$

If (*sufficient condition*), however, this Hessian is positive definite on $\mathcal{C}$, then $(0,0)$ is a (strict) local minimiser.
Since c_1 is the only active constraint and $\lambda_1 > 0$, we find that

$$\mathcal{C} = \big\{(w_1,w_2) \in \mathbb{R}^2 : \nabla c_1(0,0)^\top w = 0\big\} = \big\{(w_1,0) \in \mathbb{R}^2 : w_1 \in \mathbb{R}\big\}.$$

Therefore, with $w = (w_1,0) \in \mathcal{C}$,

$$w^\top \nabla^2_{(x,y)}\mathcal{L}(0,0,2,0) w = -2(1-2C)w_1^2,$$

which is nonnegative if and only if $C \geq 1/2$, and strictly positive for all $w \in \mathcal{C} \setminus \{0\}$ if and only if $C > 1/2$. Thus $(0,0)$ is a (strict) local minimum whenever $C > 1/2$, but cannot be a minimiser if $0 < C < 1/2$. It remains to examine $C = 1/2$. To this end, we consider, for example, points (x,y) approaching $(0,0)$ along $c_1(x,y) = 0$, that is, points for which $y = x^2/2 \to 0$. This yields

$$f\big(x, \tfrac{1}{2}x^2\big) = -x^2 - \big(\tfrac{1}{2}x^2 - 1\big)^2 = -\tfrac{1}{4}x^4 - 1,$$

which is strictly less than $f(0,0) = -1$ for all $x \neq 0$. In particular, $(0,0)$ is not a local minimiser when $C = 1/2$.

3 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 $c(x,y) = xy - 1$. By completing the square, we get that

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

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: Constructively, 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

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

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 $(x, y) = (\pm 1, \pm 1)$.

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 optimization problem

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

a) Formulate the KKT-conditions for this problem and show that the unique solution is given by

$$x^* = A^T(AA^T)^{-1}b.$$

Solution: We are now considering the problem

$$\min_{x\in\mathbb{R}^n} f(x), \text{ s.t. } c(x) = 0,$$

where

$$f(x) = \frac{1}{2}x^T x \text{ and } c(x) = Ax - b,$$

with $b \neq 0$. The Lagrangian is

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^T x - \lambda^T(Ax - b),$$

where $\lambda \in \mathbb{R}^m$. The KKT conditions become

$$\begin{aligned}\nabla\mathcal{L}(x,\lambda) = x - A^T\lambda = 0,\\ Ax - b = 0.\end{aligned}$$

Also, since A has full rank, then the LICQ hold everywhere, meaning the KKT conditions are necessary for minimizers. We therefore look for solutions that satisfy the KKT conditions. If $\lambda = 0$, then $x = 0$ and $Ax = 0$, meaning $Ax - b \neq 0$, so we must have $\lambda \neq 0$. The first condition then gives $x = A^T\lambda$, and inserting this into the second gives $AA^T\lambda = b$. Since A has full rank, AA^T is invertible and we have $\lambda = (AA^T)^{-1}b$, meaning $x = A^T(AA^T)^{-1}b$. Also, since $\nabla^2\mathcal{L}(x,\lambda) = \nabla^2 f(x) = I$, which is positive definite, this is a minimum.

b) Formulate the quadratic penalty method for this constrained optimization problem, and show that the unique minimizer with parameter $\mu > 0$ is given by

$$x_\mu := A^T\Big(\frac{1}{\mu}\,\mathrm{Id} + AA^T\Big)^{-1}b$$

with $\mathrm{Id} \in \mathbb{R}^{m\times m}$ denoting the identity matrix.

Solution: The quadratic penalty method considers the unconstrained optimization of

$$g(x) = f(x) + \frac{\mu}{2}c(x)^T c(x),$$

which in our case becomes

$$g(x) = \frac{1}{2}x^T x + \frac{\mu}{2}(Ax - b)^T(Ax - b).$$

Taking the gradient of this, we get

$$\begin{aligned}\nabla g(x) = x + \mu(A^T Ax - A^T b) = 0 \\ \Rightarrow \left(\frac{1}{\mu} I + A^T A\right) x = A^T b \\ \Rightarrow x = \left(\frac{1}{\mu} I + A^T A\right)^{-1} A^T b.\end{aligned}$$

This is, however, not the expression we were looking for. We can easily see that

$$A^T \left(\frac{1}{\mu} I + AA^T\right) = \left(\frac{1}{\mu} I + A^T A\right) A^T.$$

Multiplying both sides from the left by $\left(\frac{1}{\mu} I + A^T A\right)^{-1}$ and from the right by $\left(\frac{1}{\mu} I + AA^T\right)^{-1}$, we see that

$$\left(\frac{1}{\mu} I + A^T A\right)^{-1} A^T = A^T \left(\frac{1}{\mu} I + AA^T\right)^{-1},$$

meaning that we get

$$x = A^T \left(\frac{1}{\mu} I + AA^T\right)^{-1} b.$$

Another way of arriving at the desired expression is by use of the singular value decomposition of A, writing $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices ($U^{-1} = U^T$ and $V^{-1} = V^T$) and $\Sigma \in \mathbb{R}^{m\times n}$ is the matrix containing the singular values of A along its diagonal. The singular values are all positive. We will write $I_{r\times r}$ for an $r \times r$ identity matrix. Now, we observe

that

$$\begin{aligned}
\left(\frac{1}{\mu}I_{n\times n}+A^TA\right)^{-1}A^T &= \left(\frac{1}{\mu}I_{n\times n}+(U\Sigma V^T)^TU\Sigma V^T\right)^{-1}(U\Sigma V^T)^T\\
&= \left(\frac{1}{\mu}I_{n\times n}+V\Sigma^TU^TU\Sigma V^T\right)^{-1}V\Sigma^TU^T\\
&= \left(\frac{1}{\mu}I_{n\times n}+V\Sigma^T\Sigma V^T\right)^{-1}V\Sigma^TU^T\\
&= \left(V\left(\frac{1}{\mu}I_{n\times n}+\Sigma^T\Sigma\right)V^T\right)^{-1}V\Sigma^TU^T\\
&= V\left(\frac{1}{\mu}I_{n\times n}+\Sigma^T\Sigma\right)^{-1}V^TV\Sigma^TU^T\\
&= V\left(\frac{1}{\mu}I_{n\times n}+\Sigma^T\Sigma\right)^{-1}\Sigma^TU^T\\
&= V\Sigma\left(\frac{1}{\mu}I_{m\times m}+\Sigma\Sigma^T\right)^{-1}U^T\\
&= V\Sigma U^TU\left(\frac{1}{\mu}I_{m\times m}+\Sigma\Sigma^T\right)^{-1}U^T\\
&= A^T\left(\frac{1}{\mu}I_{m\times m}+U\Sigma\Sigma^TU^T\right)^{-1}\\
&= A^T\left(\frac{1}{\mu}I_{m\times m}+U\Sigma V^TV\Sigma^TU^T\right)^{-1}\\
&= A^T\left(\frac{1}{\mu}I_{m\times m}+AA^T\right)^{-1}.
\end{aligned}$$

Thereby, we have $x_\mu = A^T\left(\frac{1}{\mu}I_{m\times m}+AA^T\right)^{-1}b$. The fact that

$$\left(\frac{1}{\mu}I_{n\times n}+\Sigma^T\Sigma\right)^{-1}\Sigma^T = \Sigma\left(\frac{1}{\mu}I_{m\times m}+\Sigma\Sigma^T\right)^{-1}$$

can be checked by writing the product componentwise.

c) Now consider the optimization problem

$$\frac{1}{2}\|x\|^2 \to \min \qquad \text{subject to} \qquad \frac{1}{2}\|Ax-b\|^2 \le \varepsilon$$

for some $\varepsilon > 0$, and denote its solution by $\hat{x}_\varepsilon$. Show that either $\frac{1}{2}\|b\|^2 \le \varepsilon$ (in which case $\hat{x}_\varepsilon = 0$), or there exists $\mu > 0$ such that $\hat{x}_\varepsilon = x_\mu$.

Solution: We now consider the problem

$$\min_{x\in\mathbb{R}^n} f(x), \text{ s.t. } c(x) \ge 0,$$

where

$$f(x) = \frac{1}{2}x^Tx \text{ and } c(x) = \epsilon - \frac{1}{2}||Ax-b||^2,$$

The KKT conditions for this problem are

$$\begin{aligned}\nabla\mathcal{L}(x,\lambda) = x + \lambda(A^TAx - A^Tb) &= 0\\ \lambda(\epsilon - \frac{1}{2}||Ax-b||^2) &= 0\\ \epsilon - \frac{1}{2}||Ax-b||^2 &\geq 0\\ \lambda &\geq 0.\end{aligned}$$

With $\lambda = 0$, we get $x = 0$. For the third condition to hold, we must have $\epsilon \geq ||b||^2/2$. This is then a valid KKT point. Also, we have $\nabla^2\mathcal{L}(x,0) = I$, which is positive definite, so it is a minimum.
If $\lambda \neq 0$, we get, as in the previous exercise, that

$$\hat{x}_e = A^T\left(\frac{1}{\lambda}I_{m\times m} + AA^T\right)^{-1}b.$$

Here, λ must satisfy the condition that $\lambda > 0$ and λ must solve

$$\epsilon - \frac{1}{2}\left|\left|\left(AA^T\left(\frac{1}{\lambda}I_{m\times m} + AA^T\right)^{-1} - I_{m\times m}\right)b\right|\right|^2 = 0.$$

We can show that such a λ exists; since f is coercive and Ω is bounded and closed, there must exist a global minimizer. Since the LICQ holds, the KKT conditions are necessary for a minimum, and since, if $\epsilon < \frac{1}{2}||b||^2$, our candidate $\hat{x}_e$ is the only KKT point, it must be the global minimum, and thereby have a λ satisfying the above conditions. Thus, by taking $\mu = \lambda$, we get $\hat{x}_\epsilon = x_\mu$.