Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimization Theory Spring 2015

Exercise set 8 solution

$\boxed{1}$ **a)** We start by writing the problem in the form

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y), \text{ s.t. } c(x,y)=0,$$

where

$$f(x,y)=\frac{1}{2}(x^2+y^2) \text{ and } c(x,y)=xy-1.$$

We then form the Lagrangian

$$\mathcal{L}(x,y,\lambda)=f(x,y)-\lambda c(x,y),$$

and find its gradient

$$\nabla\mathcal{L}(x,y,\lambda)=\begin{bmatrix} x-\lambda y \\ y-\lambda x \end{bmatrix}.$$

We may note that $\nabla c(x,y)\neq 0$ for all $(x,y)\neq(0,0)$, meaning the LICQ holds for all feasible (x,y). We can therefore use the first order KKT conditions to identify candidates for extrema. Solving $\nabla\mathcal{L}(x,y,\lambda)=0$, we find

$$x=\lambda y \text{ and } y(1-\lambda^2)=0.$$

Thus, we may have either $\lambda=\pm 1$ or $y=0$. But, if $y=0$, then $c(x,y)=1\neq 0$, so this possibility is excluded. Also, taking $\lambda=-1$ yields imaginary values for x and y, which are disregarded. Therefore, we take $\lambda=1$ and continue by observing that

$$c(x,y)=c(y,y)=y^2-1=0\Rightarrow y=\pm 1.$$

Thereby, we have the two solutions $(x,y)=(\pm 1,\pm 1)$, both of which has Lagrange multiplier $\lambda=1$. To verify that they are indeed minima, we check the second order sufficient conditions. We have that

$$\nabla c(x,y)=\begin{bmatrix} y \\ x \end{bmatrix},$$

so the LICQ holds at both points and in addition, for any w in the critical cone at $(\pm 1,\pm 1)$, we have $w=[\gamma,-\gamma]^T$, $\gamma\in\mathbb{R}$. At the two points we then get

$$w^T\nabla^2\mathcal{L}(x,y,\lambda)w=\begin{bmatrix}\gamma & -\gamma\end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \gamma \\ -\gamma \end{bmatrix}=4\gamma^2\geq 0,$$

with equality if and only if $\gamma = 0$, i.e. $w = 0$. Thus, $(x, y) = (\pm 1, \pm 1)$ are indeed minima.

Alternatively, we could argue that since f is coercive and Ω is closed, then there exist minimizers to the problem, and since the points are KKT points where the LICQ holds, these are the only two candidates for minimizers. Then, since f takes the same value at both points, both points must be global minimizers.

b) The quadratic penalty method considers the unconstrained minimization of the objective function

$$g(x, y) = f(x, y) + \frac{\mu}{2} c(x, y)^2 = \frac{1}{2}(x^2 + y^2) + \frac{\mu}{2}(xy - 1)^2.$$

We find that

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

and solving $\nabla g(x, y) = 0$ gives (from the first component)

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

Inserting this into the second equation, we get

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

One solution of this is $y = 0$, giving $(x, y) = (0, 0)$. If $y \neq 0$, we can simplify the equation to

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

with solutions

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

which exist as long as $\mu \geq 1$. From here, one can check that this also gives

$$x = \pm \sqrt{1 - \frac{1}{\mu}}.$$

It can be checked that these points are minimizers as long as they exist. The solution $(x, y) = (0, 0)$ is a minimizer while $\mu \leq 1$, but becomes a maximizer when $\mu > 1$. Also, we can see that as $\mu \to \infty$, $(x, y) \to (\pm 1, \pm 1)$.

c) 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 $\lambda + \mu \leq -1$.

$\boxed{2}$ **a)** 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 now

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

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, 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) 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

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

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) We now consider the problem

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

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 $\hat{x}_\epsilon = x_\mu$.

3 Take, for example, the minimization of the function $f(x, y) = -x^6 - y^6$ on the unit circle, i.e. with $c(x, y) = 1 - x^2 - y^2$. We then have

$$\mathcal{L}_A(x, y) = -x^6 - y^6 - \lambda(1 - x^2 - y^2) + \mu(1 - x^2 - y^2)^2,$$

which clearly is not bounded from below.

4 We are considering the problem

$$\min_{(x,y)\in\mathbb{R}^2} f(x, y), \text{ s.t. } c(x, y) = 0,$$

where

$$f(x, y) = \frac{1}{2}(x^2 + y^2) \text{ and } c(x, y) = 1 - x.$$

We can clearly see that the minimum is found at $(x, y) = (1, 0)$. The penalty function is

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

We split the minimization problems up into the three cases $x = 1, x < 1$ and $x > 1$. When $x = 1$, we have $\Phi_1(x, y; \mu) = \frac{1}{2}(1 + y^2)$, for which a minimum is attained at $y = 0$, with the value $\Phi_1(1, 0; \mu) = \frac{1}{2}$.

For $x < 1$, we find the minimizer by solving $\nabla\Phi_1(x, y; \mu) = 0$, yielding $(x, y) = (\mu, 0)$, and with $\Phi_1(\mu, 0; \mu) = \mu(1 - \frac{\mu}{2})$. Note that this minimizer exists only when $\mu < 1$.

For $x > 1$, we find the minimizer by solving $\nabla\Phi_1(x, y; \mu) = 0$, yielding $(x, y) = (-\mu, 0)$, and with $\Phi_1(-\mu, 0; \mu) = -\mu(1 + \frac{\mu}{2})$. However, we see that this minimizer exists only when $\mu < -1$, and as such can be disregarded as we are only looking at positive values of μ.

Comparing the minimizers $(\mu, 0)$ and $(1, 0)$, we see that since $\mu(1 - \frac{\mu}{2}) < \frac{1}{2}$ when $\mu < 1$, the global minimizer of Φ_1 will be $(\mu, 0)$ until $\mu = 1$, when the minimizer becomes $(1, 0)$, the minimizer of the constrained problem. This is in accordance with theorem 17.3 in N&W, since one can check that the Lagrange multiplier for the constrained problem is $\lambda = 1$.

$\boxed{5}$ For clarity, define $h = g \circ f$. That is, let $h(x) = g(f(x))$. Then, we have

$$h(\lambda x + (1 - \lambda)y) = g(f(\lambda x + (1 - \lambda)y)).$$

Since f is convex, we know that $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$. Moreover, since g is monotone increasing, we therefore know that

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

Since g is also convex, we have

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

In total, we see that

$$h(\lambda x + (1 - \lambda)y) < \lambda g(f(x)) + (1 - \lambda)g(f(y)) = \lambda h(x) + (1 - \lambda)h(y),$$

meaning $h = g \circ f$ is convex.

$\boxed{6}$ We wish to prove that the function

$$f(x) = \begin{cases} +\infty, & x < 0 \\ -\sqrt{x}, & x \geq 0 \end{cases}$$

is convex. First, assume that either x or y is negative. Then, we clearly have

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

Next, if both x and y are strictly positive, we observe that the function $g(x) = -\sqrt{x}$ is convex; its double derivative is strictly positive for positive x. Therefore, we have

$$f(\lambda x + (1 - \lambda)y) = -\sqrt{\lambda x + (1 - \lambda)y} \leq -\lambda\sqrt{x} - (1 - \lambda)\sqrt{y} = \lambda f(x) + (1 - \lambda)f(y),$$

Furthermore, if $x = 0$ and $y > 0$, we have

$$f(\lambda x + (1 - \lambda)y) = -\sqrt{(1 - \lambda)}\sqrt{y} \leq -(1 - \lambda)\sqrt{y} = \lambda f(x) + (1 - \lambda)f(y),$$

since $\sqrt{(1 - \lambda)} \geq (1 - \lambda)$ for $\lambda \in [0, 1]$. The same argument holds for $x > 0$ and $y = 0$. Finally, if both $x = 0$ and $y = 0$, convexity holds trivially. Thereby, we have that for all $x \in \mathbb{R}$, $y \in \mathbb{R}$ and $\lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$, so f is convex.

To compute the subdifferential $\partial f(x)$ for every $x \geq 0$, we first observe that $f(x)$ is differentiable at all $x > 0$, so there we simply have

$$\partial f(x) = \left\{ \frac{\mathrm{d}f}{\mathrm{d}x}(x) \right\} = \left\{ -\frac{1}{2\sqrt{x}} \right\}.$$

At $x = 0$, the matter is slightly more complicated since f is not differentiable. We search for all subgradients of f at 0, i.e. all $\xi \in \mathbb{R}$ such that

$$f(y) \geq f(0) + \xi(y - 0) \quad \forall y \in \mathbb{R},$$

or

$$f(y) \geq \xi y \quad \forall y \in \mathbb{R}.$$

If $y < 0$, then $f(y) = +\infty$, and the inequality holds for all ξ. However, if $y > 0$, then we require

$$-\sqrt{y} \geq \xi y,$$

or

$$\xi \leq -\frac{1}{\sqrt{y}}, \quad y \in (0, +\infty).$$

Such a number does not exist, since the left hand side goes to $-\infty$ as $y \to 0$. Therefore, the subdifferential at $x = 0$ is

$$\partial f(0) = \emptyset.$$

Alternatively, we could have argued that since the subdifferential is monotone, we would have the condition that if $\xi \in \partial f(0)$, then

$$\xi \leq -\frac{1}{2\sqrt{x}} \quad \forall x > 0,$$

which is impossible.

$\boxed{7}$ We wish to find the convex conjugate of the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \mathrm{e}^x$. By definition we have

$$f^*(\xi) = \sup_{x \in \mathbb{R}} \xi x - f(x),$$

meaning we have to maximize the function $g(x) = \xi x - \mathrm{e}^x$ for each ξ. We find that

$$g'(x) = \xi - \mathrm{e}^x \text{ and } g''(x) = -\mathrm{e}^x,$$

so $g(x)$ is concave, meaning any x s.t. $g'(x) = 0$ is a maximizer. Solving $g'(x) = 0$, we get $x = \ln(\xi)$, which works if $\xi > 0$. Otherwise, we have $g'(x) < 0$ for all x, meaning $g(x)$ is monotone decreasing such that $\sup_x g(x)$ is obtained in the limit $x \to -\infty$. Thus, we get

$$f^*(\xi) = \begin{cases} \xi \ln \xi - \xi, & \xi > 0, \\ 0, & \xi = 0, \\ +\infty, & \xi < 0. \end{cases}$$

$\boxed{8}$ We wish to prove that the function

$$f(x,y) = \begin{cases} x^2/y, & y > 0 \\ 0, & x = y = 0 \\ +\infty, & \text{otherwise} \end{cases}$$

is convex. That is, given two points (x_1, y_1) and (x_2, y_2), we want to show that

$$f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq \lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2) \tag{1}$$

when $\lambda \in (0,1)$. First, if $y_1 \leq 0, x_1 \neq 0$ or $y_2 \leq 0, x_2 \neq 0$, (1) holds since the right hand side is infinite, similarly as in exercise 6.

If $y_1 > 0$ and $y_2 > 0$, (1) will also hold since the Hessian of f is defined and positive semi-definite for all $y > 0$. This can be seen as follows:

$$\nabla^2 f(x,y) = \begin{bmatrix} \dfrac{2}{y} & -\dfrac{2x}{y^2} \\ -\dfrac{2x}{y^2} & \dfrac{2x^2}{y^3} \end{bmatrix}.$$

Its eigenvalues γ are given by the characteristic equation

$$\left(\frac{2}{y} - \gamma\right)\left(\frac{2x^2}{y^3} - \gamma\right) - \frac{4x^2}{y^4} = 0,$$

which can be solved to find

$$\gamma_1 = 0 \qquad \gamma_2 = \frac{2x^2 + 2y^2}{y}.$$

Since γ_1 and γ_2 are non-negative, the Hessian is positive semi-definite. Finally, if $(x_1, y_1) = (0,0)$, and $y_2 \leq 0, x_2 \neq 0$, we have

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) &= f((1-\lambda)x_2, (1-\lambda)y_2) \\ &= (1-\lambda)\frac{x_2^2}{y_2} \\ &= \lambda f(x_1, y_1) + (1-\lambda) f(x_2, y_2), \end{aligned}$$

and so (1) holds with equality. The same applies to the case $(x_2, y_2) = (0,0)$, and $y_1 \leq 0, x_1 \neq 0$. If $(x_1, y_1) = (x_2, y_2) = (0,0)$, (1) holds trivially. Thus, we have shown that f is convex.

The subdifferential $\partial f(x,y)$ can be easily computed when $y > 0$; we then have

$$\partial f(x,y) = \{\nabla f(x,y)\} = \left\{ \begin{bmatrix} \dfrac{2x}{y} \\ -\dfrac{x^2}{y^2} \end{bmatrix} \right\}.$$

At the point $(0,0)$, we once more need to be careful. We look for $\xi \in \mathbb{R}^2$ such that

$$f(x,y) \geq f(0,0) + \xi^T \begin{bmatrix} x \\ y \end{bmatrix} \qquad \forall\, (x,y) \in \mathbb{R}^2.$$

If $y < 0$, this clearly holds since then we have $f(x, y) = +\infty$. If $y > 0$, we look for components ξ_1 and ξ_2 such that

$$\xi_1 x + \xi_2 y \leq \frac{x^2}{y} \qquad \forall x \in \mathbb{R}, y \in (0, \infty). \tag{2}$$

One way of arriving at the answer is by rearranging (2) to:

$$x^2 - \xi_1 xy - \xi_2 y^2 \geq 0,$$

and considering this as a polynomial in x for some arbitrary value $y > 0$. Since the polynomial has a positive leading term, this is equivalent to requiring that the polynomial have at most one real root. This is achieved if the discriminant is non-positive, i.e.

$$\xi_1^2 y^2 + 4\xi_2 y^2 \leq 0 \quad \forall y > 0,$$

implying that $\xi_1^2 \leq -4\xi_2$, which can only hold if $\xi_2 \leq 0$. Thereby, the subdifferential at 0 is

$$\partial f(0, 0) = \{(\xi_1, \xi_2) \mid \xi_1^2 \leq -4\xi_2 \text{ and } \xi_2 \leq 0.\}.$$

Another way of arriving at the same conclusion is as follows: we may begin by observing that by taking $x = 0$, we get

$$\xi_2 y \leq 0, \qquad \forall y \in (0, \infty),$$

meaning we must have $\xi_2 \leq 0$. First, assume that $\xi_2 = 0$. Then, we require

$$\xi_1 x \leq \frac{x^2}{y} \qquad \forall x \in \mathbb{R}, y \in (0, \infty).$$

The only possibility for this to hold is taking $\xi_1 = 0$. Otherwise, we could take $y = 1$ and $x = \operatorname{sgn}(\xi_1)\gamma$ with $\gamma < |\xi_1|$ as a counterexample. Thus, the choice $(\xi_1, \xi_2) = (0, 0)$ is a possibility. Next, consider the choice of $\xi_2 < 0$. We may then write $\xi_2 = -\alpha^2$, $\alpha \in \mathbb{R}\backslash\{0\}$. The inequality (2) is then equivalent to:

$$\xi_1 xy - \alpha^2 y^2 - x^2 \leq 0 \qquad \forall x \in \mathbb{R}, y \in (0, \infty),$$

which (by completing the square) is equivalent to:

$$-\left(\alpha y - \frac{\xi_1 x}{2\alpha}\right)^2 - x^2\left(1 - \frac{\xi_1^2}{4\alpha^2}\right) \leq 0 \qquad \forall x \in \mathbb{R}, y \in (0, \infty).$$

Changing sides in this, we get the equivalent condition

$$x^2\left(\frac{\xi_1^2}{4\alpha^2} - 1\right) \leq \left(\alpha y - \frac{\xi_1 x}{2\alpha}\right)^2 \qquad \forall x \in \mathbb{R}, y \in (0, \infty).$$

This will hold if the left hand side is negative, i.e.

$$\left(\frac{\xi_1^2}{4\alpha^2} - 1\right) \leq 0,$$

meaning

$$\xi_1^2 \leq -4\xi_2.$$

Otherwise, we may choose $x = \frac{2\alpha^2 y}{\xi_1}$, making the right hand side zero but the left hand side positive. To summarize, the relation (2) holds if

$$\xi_1^2 \leq -4\xi_2 \quad \text{and} \quad \xi_2 \leq 0.$$

9 We wish to prove that the function

$$f(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq 1 \\ |x| - \frac{1}{2}, & |x| > 1 \end{cases}$$

is convex. It is continuously differentiable, with

$$f'(x) = \begin{cases} -1, & x < -1 \\ x, & |x| \leq 1 \\ 1, & x > 1. \end{cases}$$

Since $f'(x)$ is monotone increasing, $f(x)$ is convex. We find its conjugate f^* by the definition

$$f^*(\xi) = \sup_{x \in \mathbb{R}} \xi x - f(x)$$

To find this, we define $g(x) = \xi x - f(x)$ and find its supremum through considering its derivative,

$$g'(x) = \begin{cases} \xi + 1, & x < -1 \\ \xi - x, & |x| \leq 1 \\ \xi - 1, & x > 1. \end{cases}$$

We see that if $\xi > 1$, then $g'(x) > 0$, and therefore $\sup g(x) = \infty$. Also, if $\xi < -1$, we have $g'(x) < 0$, meaning $\sup g(x) = \infty$ (approached in the limit $x \to -\infty$). Otherwise, if $|\xi| \leq 1$, we have $g'(x) = 0$ at $x = \xi$, with function value $g(\xi) = \frac{\xi^2}{2}$. These are maxima, since $g''(x) = -1$ at these points. In total, we get

$$f^*(\xi) = \begin{cases} \frac{\xi^2}{2}, & |\xi| \leq 1 \\ +\infty, & |\xi| > 1. \end{cases}$$

10 We wish to prove that if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is strictly convex and $x \neq y \in \mathbb{R}^n$, then $\partial f(x) \cap \partial f(y) = \emptyset$. This can be done through proof by contradiction. Assume that $\partial f(x) \cap \partial f(y) \neq \emptyset$. Then, there exists a $\xi \in \mathbb{R}^n$ such that

$$f(z) \geq f(x) + \xi^T(z - x), \quad \forall z \in \mathbb{R}^n \tag{3}$$

$$f(z) \geq f(y) + \xi^T(z - y), \quad \forall z \in \mathbb{R}^n. \tag{4}$$

Furthermore, since f is strictly convex, we know that

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y). \quad \forall \lambda \in (0, 1)$$

Inserting (3) and (4) into this inequality, we find

$$f(\lambda x + (1 - \lambda)y) < f(z) + \lambda \xi^T(z - x) + (1 - \lambda)\xi^T(z - y),$$

which should hold for all $z \in \mathbb{R}^n$. But now, if we take $z = \lambda x + (1 - \lambda)y$ for some $\lambda \in (0, 1)$ and clear up the terms, we come to the conclusion that

$$f(\lambda x + (1 - \lambda)y) < f(\lambda x + (1 - \lambda)y),$$

which is a contradiction. Therefore, we cannot have $\partial f(x) \cap \partial f(y) \neq \emptyset$, meaning $\partial f(x) \cap \partial f(y) = \emptyset$.

$\boxed{11}$ **a)** We wish to find the function $y(x)$ minimizing the functional

$$S(y) = 2\pi \int_0^l y(x)\sqrt{1+y'(x)^2}\mathrm{d}x.$$

To this end, we identify the function

$$f(x, y(x), y'(x)) = y(x)\sqrt{1+y'(x)^2},$$

and see that from Theorem 4 in the note on the calculus of variations that f must satisfy the Euler-Lagrange equation

$$\frac{\partial}{\partial y} f(x, y(x), y'(x)) = \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial}{\partial y'} f(x, y(x), y'(x)) \tag{5}$$

with the boundary condition $y(0) = a$, $y(l) = b$. We can see that

$$\frac{\partial}{\partial y} f(x, y(x), y'(x)) = \sqrt{1+y'(x)^2}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial}{\partial y'} f(x, y(x), y'(x)) = \frac{y'(x)^2 + y(x)y''(x)}{\sqrt{1+y'(x)^2}} - \frac{y(x)y'(x)^2y''(x)}{(1+y'(x)^2)^{3/2}}.$$

Inserting these two expressions into (5) and collecting terms gives the differential equation

$$\frac{1+y'(x)^2 - y(x)y''(x)}{(1+y'(x)^2)^{3/2}} = 0,$$

which, since the numerator is strictly positive, is equivalent to the differential equation

$$1 + y'(x)^2 - y(x)y''(x) = 0. \tag{6}$$

This is the Euler-Lagrange equation for the variational functional $S(y)$.

b) With the assumption that

$$y(x) = A\cosh\left(\frac{x-B}{A}\right),$$

we have

$$y'(x) = \sinh\left(\frac{x-B}{A}\right),$$
$$y''(x) = \frac{1}{A}\cosh\left(\frac{x-B}{A}\right).$$

Inserting this into (6), we find

$$1 + y'(x)^2 - y(x)y''(x) = 1 + \sinh^2\left(\frac{x-B}{A}\right) - \cosh^2\left(\frac{x-B}{A}\right) = 0,$$

since $\cosh^2(z) - \sinh^2(z) = 1$. The constants A and B can be determined by solving the set of equations

$$y(0) = A\cosh\left(\frac{-B}{A}\right) = a$$
$$y(l) = A\cosh\left(\frac{l-B}{A}\right) = b.$$

$\boxed{12}$ In both a) and b), we wish to find the function $y(x)$ minimizing the functional

$$S(y) = \frac{1}{2}\int_0^1 y(x)^2 + y'(x)^2 \mathrm{d}x.$$

To this end, we identify the function

$$f(x, y(x), y'(x)) = y(x)^2 + y'(x)^2,$$

and find that y must satisfy the Euler-Lagrange equation

$$y(x) - y''(x) = 0,$$

meaning we have

$$y(x) = A\mathrm{e}^x + B\mathrm{e}^{-x},$$

which is equivalent to

$$y(x) = A\cosh(x) + B\sinh(x).$$

The latter form of y is easier to work with for our purposes. The remaining task is to find the constants A and B, which must be determined through boundary conditions.

a) Here, we want y to satisfy the boundary conditions

$$y(0) = A = 1,$$
$$y(1) = A\cosh(1) + B\sinh(1) = 0.$$

This can easily be solved to yield $A = 1$, $B = -1/\tanh(1)$, meaning we have

$$y(x) = \cosh(x) - \frac{\sinh(x)}{\tanh(1)}.$$

b) In this point, we want $y(0) = 1$, but with the natural boundary condition (as described in section 6 in the note on variational calculus) at $x = 1$:

$$\frac{\partial}{\partial y'} f(1, y(1), y'(1)) = 2y'(1) = 0.$$

Thereby, we get the boundary condition equations

$$y(0) = A = 1,$$
$$y'(1) = A\sinh(1) + B\cosh(1) = 0,$$

which can be solved to yield $A = 1$, $B = -\tanh(1)$, meaning we have

$$y(x) = \cosh(x) - \tanh(1)\sinh(x).$$

$\boxed{13}$ We wish to find the function $y(x)$ minimizing the functional

$$S(y) = \int_0^2 \mathrm{e}^{-x} y'(x)^2 - y(x)\mathrm{d}x,$$

with the constraints $y(0) = 1$ and $y(2) = 0$. To this end, we identify the function

$$f(x, y(x), y'(x)) = \mathrm{e}^{-x} y'(x)^2 - y(x),$$

and find that y must satisfy the Euler-Lagrange equation

$$-1 + 2\mathrm{e}^{-x} y'(x) - 2\mathrm{e}^{-x} y''(x) = 0,$$

which is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{-x} y'(x)) = -\frac{1}{2}.$$

From this, we get

$$y'(x) = -\frac{x\mathrm{e}^x}{2} + C\mathrm{e}^x,$$

which can be solved to yield

$$y(x) = -\frac{x\mathrm{e}^x}{2} + \left(C + \frac{1}{2}\right)\frac{\mathrm{e}^x}{2} + D.$$

Imposing boundary conditions, we find

$$y(0) = C + \frac{1}{2} + D = 1 \Rightarrow C = \frac{1}{2} - D$$

and so

$$y(2) = -\mathrm{e}^2 + \mathrm{e}^2 - D\mathrm{e}^2 + D = 0 \Rightarrow D = 0 \Rightarrow C = \frac{1}{2},$$

which yields

$$y(x) = -\frac{x\mathrm{e}^x}{2} + \mathrm{e}^x.$$