Exercises for Lecture Course on Numerical Optimization (NUMOPT)
Albert-Ludwigs-Universität Freiburg – Summer Semester 2025

Exercise 2: Convexity, Duality and Fitting problems

Prof. Dr. Moritz Diehl and tutors.

The solutions for these exercises will be given and discussed during the exercise session on May 13th. To receive feedback on your solutions, please hand it in during the exercise session on May 13th, or by e-mail to leo.simpson@imtek.uni-freiburg.de before the same date.

I Convex optimization

In this part we learn how to recognize convex sets and functions. Moreover we revisit the hanging chain problem from the previous exercise sheet and add different types of constraints.

I.1 Convex sets

Determine whether the following sets are convex or not. Justify your answers.

1. A halfspace, i.e., a set of the form:

$$\Omega = \{x \in \mathbb{R}^n | a^\top x \leq b\}$$

for some $a \in \mathbb{R}^n$, and $b \in \mathbb{R}$.

Solution: A halfspace is convex.
To show that a set is convex, we need to show that for any $x, y \in \Omega$ and $t \in [0, 1]$, the point $z = (1 - t)x + ty$ is also in Ω.

$$z = (1-t)x + ty \Rightarrow a^\top z = (1-t)a^\top x + ta^\top y \leq (1-t)b + tb = b$$

Thus, $z \in \Omega$.

Alternatively, the function $f(x) = a^\top x - b$ is linear and therefore convex. Then, the set $\Omega = \{x \in \mathbb{R}^n \mid f(x) \leq 0\}$ is convex.

2. A wedge, i.e., a set of the form

$$\Omega = \{x \in \mathbb{R}^n \mid a_1^\top x \leq b_1, a_2^\top x \leq b_2\}$$

for some $a_1, a_2 \in \mathbb{R}^n$, and $b_1, b_2 \in \mathbb{R}$.

Solution: Each halfspace is convex. Hence, a wedge, which is the intersection of two halfspaces, is convex.

Alternative: show set definition also holds for $z = (1 - t)x + ty$ with x, y elements of the set.

3. The set of points closer to a given point than to a given set:

$$\Omega = \{x \in \mathbb{R}^n \mid \|x - x_0\|_2 \leq \|x - y\|_2 \ \forall y \in \mathcal{S}\}$$

where $x_0 \in \mathbb{R}^n$ is a vector and $\mathcal{S} \subseteq \mathbb{R}^n$ is a set of points.

Solution: The set can be equivalently written as the intersection

$$\Omega = \bigcap_{y \in \mathcal{S}} \{x \in \mathbb{R}^n \mid \|x - x_0\|_2 \leq \|x - y\|_2\} = \bigcap_{y \in \mathcal{S}} \Omega_y.$$

Now each Ω_y defines a halfspace, which is convex. Then Ω is an intersection of convex sets and therefore convex itself.
To see each Ω_y defines a halfspace:

$$\begin{aligned}
& \|x - x_0\|^2 \leq \|x - y\|^2, \\
\iff & \|x\|^2 + 2x^\top x_0 + \|x_0\|^2 \leq \|x\|^2 + 2x^\top y + \|y\|^2 \\
\iff & \underbrace{(y - x_0)^\top}_{a_y^\top} x \leq \underbrace{\frac{1}{2}\left(\|y\|^2 - \|x_0\|^2\right)}_{b_y} \\
\iff & a_y^\top x \leq b_y
\end{aligned}$$

4. The set of points closer to one set than to another:

$$\Omega = \{x \in \mathbb{R}^n \mid \|x - z\|_2 \leq \|x - y\|_2 \ \forall y \in \mathcal{S}, \forall z \in \mathcal{T}\}$$

where $x_0 \in \mathbb{R}^n$ is a vector and $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$ are sets of points.

Solution: Not convex.
Counter example: $n = 1$, $\mathcal{S} = \{-1, 1\}$, $\mathcal{T} = \{0\} \Rightarrow \mathcal{C} = \{x \in \mathbb{R} \mid x \leq -\frac{1}{2} \text{ or } x \geq \frac{1}{2}\}$

I.2 Convex functions

Determine whether the following functions are convex or not. Justify your answers.

1. The function $f(x) = -\log(x)$ on $\mathbb{R}_{++}$.

Solution: This is a C^2 function, so it is convex if and only if its Hessian (=second derivative) is positive semidefinite on the domain ($\mathbb{R}^2_{++}$).

$$f''(x) = \frac{1}{x^2} \geq 0$$

Hence, this function is convex.

2. The function $f(x) = x^3$ on $\mathbb{R}$.

Solution: Not convex because $f''(x) = 6x$, which is negative for $x < 0$.

3. The function $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$.

Solution: This is a C^2 function, so it is convex if and only if its Hessian is positive semidefinite on the domain ($\mathbb{R}^2_{++}$).
Its Hessian is given by:

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix} = \frac{1}{x_1^3 x_2^3} \begin{bmatrix} 2x_2^2 & x_1 x_2 \\ x_1 x_2 & 2x_1^2 \end{bmatrix}$$

Note that a matrix is semidefinite if and only all its principal minors are ≥ 0.
It appears that the matrix is $\begin{bmatrix} 2x_2^2 & x_1 x_2 \\ x_1 x_2 & 2x_1^2 \end{bmatrix}$ is positive definite (first principal minor is $2x_2^2$ and second is its determinant: $x_1^2 x_2^2$).
Conclusion: f is convex on $\mathbb{R}^2_{++}$.

4. The function $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.

Solution: The Hessian is:

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{bmatrix}$$

which is not positive semidefinite, because its determinant is negative:

$$\det \nabla^2 f(x_1, x_2) = -\frac{1}{x_2^4}$$

Conclusion: f is not convex on $\mathbb{R}^2_{++}$.

I.3 Hanging chain, revisited

Recall the hanging chain problem from the previous exercise sheet.

1. What would happen if you add, instead of the linear ground constraints, the nonlinear ground constraints $z_i \geq -0.2 + 0.1 y_i^2$, for $i = 2, \ldots, N-1$ to your problem? What type of optimization problem is the resulting problem? Is it convex?

Solution: The objective is still quadratic, but now there are quadratic constraints as well. Therefore it is no longer a QP. Instead it is a Quadratically Constrained Quadratic Program (QCQP).

For each $i = 2, \ldots, N-1$, the constraint describes a parabola opened in positive z-direction, where all points above the parabola are part of the set. This is a convex set, therefore the NLP is still convex.

Alternative: bring into standard NLP form $h(x) \geq 0$. This is a concave function, therefore the NLP is still convex (see lecture notes theorem 3.4).

2. What would happen if you add instead the nonlinear ground constraints $z_i \geq -y_i^2$, $i = 2, \ldots, N-1$? Do you expect this optimization problem to be convex?

Solution: The additional constraints are still quadratic, therefore it is still a QCQP.

The constraint describes a parabola opened in negative z-direction, where all points above the parabola are part of the set. This is not a convex set, therefore the NLP is not convex.

3. Do you expect these two problems to have several local minima? Why?

Solution: The first problem is strictly convex, so we expect only one local minima, but for the second, there might be several local minima.

4. Now solve both variations using CasADi and plot the results (both chain and constraints).

Remark: *For readibility of your code, we suggest to introduce a variable* `TYPE_OF_CONTRAINTS` *to easily switch between the different constraints that are proposed in the serie of hanging chain exercises.*

Solution: See code in `hanging_chain_rev.py`.

5. Find numerically two different local minimum for the second variation, by initializing the solver with different initial values.

Hint: *In CasADi, you can provide an initial value* `x0` *for variable* `x` *via*

```
opti.set_initial(x,x0)
```

Solution: As demonstrated in code, for the non-convex constraint $z_i \geq -y_i^2$, different initializations find different solutions (local optima).

II Minimum of a coercive function

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive when $\lim_{\|x\|\to\infty} f(x) = +\infty$, i.e. $\forall\, M,\ \exists\, R$ such that for all $\|x\| > R$, $f(x) > M$.

Prove that for a function $f : \mathbb{R}^n \to \mathbb{R}$ continuous and coercive, the unconstrained optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

always has (at least) one global minimum point.
Hint: *Use the Weierstrass Theorem.*

Solution: Apply the definition of a coercive function to $M = f(0)$. Then we know that there exists $R \in \mathbb{R}$ such that for all $x \in \mathbb{R}^n$, if $\|x\| > R$, then $f(x) > f(0)$.
Now, consider the compact set $\Omega := \{x : \|x\| \leq R\}$. Using the Weierstrass Theorem, we know that there exists a global minimizer $x^\star \in \Omega$ to the following problem:

$$\underset{x \in \Omega}{\text{minimize}} \quad f(x)$$

Now, let $x \in \mathbb{R}^n \setminus \Omega$. By construction, we have $f(x) > f(0)$. Since $0 \in \Omega$, we know that $f(x^\star) \leq f(0)$. Therefore, $f(x) > f(x^\star)$.

Conclusion: $x^\star$ is not only a minimizer of f on Ω, but also a minimizer of f on $\mathbb{R}^n$.

III Lagrange duality and dual problems

III.1 Logarithmic barrier

Consider the following *logarithmic barrier* problem,

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & c^\top x - \sum_{j=1}^{n} \log x_j \\ \text{subject to} & a^\top x = b \end{array} \tag{1}$$

where $a, c \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Remark: Problems using a logarithmic barrier as the one above will be at the core of interior point methods that we will analyze later in this course.

Remark: Even though $\log x$ is only defined for $x > 0$, for some reasons, it is not really a problem here. Consider that we extend its definition with $\log x = -\infty$ for $x \leq 0$.

1. Derive the explicit form of the dual of this problem.

Solution: The Lagrangian of the problem reads, for some multiplier $\lambda \in \mathbb{R}$:

$$\mathcal{L}(x, \lambda) := c^\top x - \sum_{j=1}^{n} \log x_j - \lambda(a^\top x - b) = (c - \lambda a)^\top x - \sum_{j=1}^{n} \log x_j + \lambda b,$$

The dual problem is:

$$\max_{\lambda \in \mathbb{R}} \underbrace{\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda)}_{=: q(\lambda)}.$$

Let us simplify the dual function $q(\lambda)$:

$$\begin{aligned} q(\lambda) &= \min_{x \in \mathbb{R}^n} \ \mathcal{L}(x, \lambda), \\ &= \lambda b + \min_{x \in \mathbb{R}^n} \ (c - \lambda a)^\top x - \sum_{j=1}^{n} \log x_j \\ &= \lambda b + \min_{x \in \mathbb{R}^n} \ \sum_{j=1}^{n} (c - \lambda a)_j x_j - \log x_j \\ &= \lambda b + \sum_{j=1}^{n} \min_{x_j \in \mathbb{R}} \ (c - \lambda a)_j x_j - \log x_j \\ &= \lambda b + \sum_{j=1}^{n} \begin{cases} 1 + \log\left((c - \lambda a)_j\right) & \text{if } (c - \lambda a)_j > 0 \\ -\infty & \text{if } (c - \lambda a)_j \leq 0 \end{cases} \\ &= \begin{cases} \lambda b + n \sum_{j=1}^{n} \log\left((c - \lambda a)_j\right) & \text{if } (c - \lambda a)_j > 0 \text{ for all } j \\ -\infty & \text{else} \end{cases} \end{aligned}$$

Therefore, we obtain the following dual problem:

$$\begin{array}{ll} \underset{\lambda \in \mathbb{R}}{\text{maximize}} & n + \lambda b + \sum_{j=1}^{n} \log(c_j - \lambda a_j) \\ \text{subject to} & c - \lambda a \geq 0 \end{array}$$

2. What connection is there between the optimal value of the dual problem, and the optimal value of the initial problem (1)?

   Hint: You are asked to find a "strong" connection.

Solution: The primal problem is convex. Furthermore, the Slatter condition holds (i.e. there exists a point where the nonlinear constraints are strictly satisfied) because there is no nonlinear inequality constraint .
Therefore, we can apply Theorem 4.5 which states that strong duality holds. Therefore, the optimal value of the dual problem is also the optimal value of the primal (1).

III.2 Linear programming

Consider the following *integer linear program* (ILP):

$$\begin{aligned} \underset{x \in \{0,1\}^n}{\text{minimize}} \quad & -c^\top x \\ \text{subject to} \quad & Ax \geq b \end{aligned} \tag{2}$$

where $c \in \mathbb{R}^n$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The optimization variables x_i are restricted to take values in $\{0, 1\}$. Solving such problems is in general a challenging task. A common practice is to reformulate the binary constraint.

1. Reformulate the mixed-integer program (2) into a nonlinear program in the standard form:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0, \\ & h(x) \geq 0 \end{aligned} \tag{3}$$

   where the functions $f(\cdot), g(\cdot)$, and $h(\cdot)$ are continuous functions that you have to provide.

   Hint: What are the $x_i \in \mathbb{R}$ such that $x_i(1 - x_i) = 0$?

Solution: The ILP (2) is equivalent to:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & -c^\top x \\ \text{subject to} \quad & Ax \geq b, \\ & x_i(1 - x_i) = 0 \quad \text{for } i = 1, \ldots, n \end{aligned}$$

2. Is this reformulation convex?

Solution: no, it has nonlinear equality constraints, hence is not convex.

3. Write the Lagrangian of the reformulation.

Solution: The Lagrangian of the problem reads, for $\lambda \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^m$:

$$\begin{aligned}\mathcal{L}(x, \lambda, \mu) &= -c^\top x - \mu^\top (Ax - b) - \sum_{i=1}^{n} \lambda_i x_i (1 - x_i) \\ &= \left(\sum_{i=1}^{n} \lambda_i x_i^2\right) - c^\top x - \lambda^\top x - \mu^\top (Ax - b) \\ &= b^\top \mu + \left(\sum_{i=1}^{n} \lambda_i x_i^2\right) - \left(c + \lambda + A^\top \mu\right)^\top x\end{aligned}$$

4. Derive the explicit form of the dual of your reformulation.

Solution: The dual function is:

$$\begin{aligned}q(\lambda, \mu) &:= \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu), \\ &= b^\top \mu + \min_{x \in \mathbb{R}^n} \left(\sum_{i=1}^{n} \lambda_i x_i^2\right) - \left(c + \lambda + A^\top \mu\right)^\top x, \\ &= b^\top \mu + \sum_{i=1}^{n} \min_{x_i \in \mathbb{R}} \left(\lambda_i x_i^2 - \left(c + \lambda + A^\top \mu\right)_i x_i\right), \\ &= b^\top \mu - \frac{1}{2} \sum_{i=1}^{n} \begin{cases} \frac{\left(c_i + \lambda_i + \left(A^\top \mu\right)_i\right)^2}{\lambda_i} & \text{if } \lambda_i > 0, \\ \lim\limits_{\lambda_i > 0, \lambda_i \to 0} \frac{\left(c_i + \lambda_i + \left(A^\top \mu\right)_i\right)^2}{\lambda_i} & \text{if } \lambda_i = 0, \\ +\infty & \text{if } \lambda_i < 0, \end{cases}\end{aligned}$$

Hence, the dual problem takes the following form:

$$\begin{aligned}\underset{\lambda \in \mathbb{R}^n, \mu \in \mathbb{R}^m}{\text{maximize}} \quad & b^\top \mu - \frac{1}{2} \sum_{i=1}^{n} \frac{\left(c_i + \lambda_i + \left(A^\top \mu\right)_i\right)^2}{\lambda_i} \\ \text{subject to} \quad & \mu \geq 0, \\ & \lambda \geq 0\end{aligned}$$

5. Is the dual problem convex?

Solution: Yes, the dual problem is always convex.

6. Using what you have found so far, propose a procedure to compute a lower bound to the ILP (2).
Hint: we do not ask you to compute this lower bound! only explain how to compute it numerically...

Solution: Solving the dual problem from the previous question provides a lower bound to the primal of the reformulation (and hence to the initial problem).

IV Fitting problems

IV.1 Affine L_2 fitting

We consider the linear regression task: we want to predict some output $y \in \mathbb{R}$ using some input $x \in \mathbb{R}$. Some input-output data $(x_i, y_i) \in \mathbb{R}^2$ are available for $i = 1, \dots, N$, and we want to use them to estimate an affine model, i.e.

$$y_i \approx a x_i + b, \tag{4}$$

where the parameters a and b have to be estimated. For this purpose, we solve the following least-square problem:

$$\underset{a,\, b \,\in\, \mathbb{R}}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^{N} (a x_i + b - y_i)^2 \tag{5}$$

1. Rewrite the least square problem (5) in matrix form:

$$\underset{a,\, b \,\in\, \mathbb{R}}{\text{minimize}} \quad \frac{1}{2} \left\| J \begin{bmatrix} a \\ b \end{bmatrix} - y \right\|_2^2 \tag{6}$$

(you have to the matrix J and the vector y).

Solution: $y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix}$ and $J = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \in \mathbb{R}^{N \times 2}$.

2. Assume that J is full column-rank. Give an explicit form of the solution $(\hat{a}, \hat{b}) \in \mathbb{R}^2$ to (5).

Hint: J is full column-rank if and only if $J^\top J$ is invertible.

Solution:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(J^\top J\right)^{-1} J^\top y$$

3. Create a Python script that generates the data points (x_i, y_i). For this, you will choose:

- $N = 30$
- $x_1, \dots, x_N$ to be N equally spaced points in the interval $[0, 5]$,
- y_i is an actually affine function of the input, but corrupted with random noise. More precisely: $y_i = 3x_i + 4 + \eta_i$, where η_i is sampled from the normal distribution $\mathcal{N}(0, 1)$.

Plot the data points.

Hint: look up the `numpy.linspace` command and the function `numpy.random.Generator.normal` in NumPy documentation

Solution: See Python code in `fitting.py`

4. Calculate the estimates $\hat{a}, \hat{b}$ in Python using what you have found in question 2. Plot the corresponding line in the same plot as the data points.

Solution: See Python code in `fitting.py`

5. Introduce 3 outliers in y by replacing arbitrary measurements with some nonsense data and plot the new fitted line in your plot. Comment the result.

Solution: See Python code in `fitting.py` The outliers damage the fit significantly.

You will need the measurements y (both with and without outliers) and the matrix J for the next task.

IV.2 Affine L_1 fitting

In this exercise, like in the previous one, we will learn an affine model from an input-output data set. However, instead of using the least squares formulation, we will try another cost function, *the least absolute deviations*:

$$\underset{a,\,b\,\in\,\mathbb{R}}{\text{minimize}} \quad \sum_{i=1}^{N} |ax_i + b - y_i| \tag{7}$$

1. Problem (7) is not differentiable. Find an (equivalent) smooth reformulation.

   <u>*Hint:*</u> *Introduce slack variables $s_1, \ldots, s_N \in \mathbb{R}$ as additional decision variables.*

   <u>*Hint:*</u> *Note that $s \geq |x|$ if and only if $s \pm x \geq 0$.*

   <u>*Hint:*</u> *The resulting problem will be a Linear Program (LP).*

   Solution:

   $$\begin{aligned} \underset{a,\,b\,\in\,\mathbb{R},\,s\,\in\,\mathbb{R}^N}{\text{minimize}} \quad & \sum_{i=1}^{N} s_i \\ \text{subject to} \quad & s_i - (ax_i + b - y_i) \geq 0, \\ & s_i + (ax_i + b - y_i) \geq 0 \end{aligned}$$

2. Solve the optimization problem you have just formulated using CasADi for the data generated in the previous exercise (both with and without outliers).

   Plot the results against those of the L2 fitting problem.

   Solution: See Python code in `fitting.py`

3. Which norm performs better? Interpret the results.

   Solution: The L1 norm is more robust against the outliers (as it does not penalize the model-measurement-mismatch quadratically). Which norm performs better depends on the context, but here it seems like we want our method to 'ignore' the outliers (the outliers seem nonsensical). That means L1 performs better.

IV.3 Regularized linear least squares

Given a matrix $J \in \mathbb{R}^{m \times n}$, a symmetric positive definite matrix $Q \succ 0$, a vector of measurements $y \in \mathbb{R}^m$ and a point, and a penalization parameter $\alpha \geq 0$, we consider the following regularized least squares problem:

$$\underset{x\,\in\,\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|y - Jx\|_2^2 + \frac{\alpha}{2} x^\top Q x \tag{8}$$

1. Does the problem (8) always have a unique solution? Justify.

   <u>*Hint:*</u> *Distinguish the casew where $\alpha = 0$ from the case where $\alpha > 0$. Also distinguish the case where J is full column rank and from the case where it is not.*

Solution:

- For $\alpha > 0$: The problem is strictly convex, and therefore has a unique solution.
- For $\alpha = 0$: if J is full column rank, the problem is still strictly convex and has a unique solution. If J is not full column rank, the problem is not strictly convex and has infinitely many solutions.

2. We denote by $x^\star(\alpha)$ the solution of the initial problem (8) for a given $\alpha > 0$. Write $x^\star(\alpha)$ explicitly.

Solution:

$$x^\star(\alpha) = \left(J^\top J + \alpha Q\right)^{-1} J^\top y$$

3. Transform the problem into the standard regularized least-squares form:

$$\underset{\tilde{x} \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\left\|y - \tilde{J}\tilde{x}\right\|_2^2 + \frac{\alpha}{2}\|\tilde{x}\|^2 \tag{9}$$

where $\tilde{J}$ should be written explicitly, and the connection between $\tilde{x}$ and x should be given.

Hint: Use matrix square-root

Solution: Since Q is a positive definite matrix, it has a positive definite square root $Q^{1/2}$, i.e. a PD matrix such that $Q^{1/2}Q^{1/2} = Q$.
Then, perform the change of variable $\tilde{x} = Q^{1/2}x$ to transform (8) into:

$$\underset{\tilde{x} \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\left\|y - JQ^{-1/2}\tilde{x}\right\|_2^2 + \frac{\alpha}{2}\|\tilde{x}\|^2$$

Finally, simply define $\tilde{J} = JQ^{-1/2}$.

4. Prove that the solution $x^\star(\alpha)$ converges to some vector $x^\star \in \mathbb{R}^n$ when $\alpha \to 0$ (even when J is not full column-rank). Express $x^\star$ explicitly using $\tilde{J}$.

Hint: Do not use the explicit formula for $x^\star(\alpha)$. Instead, use the result from last question and Lemma 6.1 from the lecture notes about matrix pseudo-inverse.

Solution: Let $\tilde{x}^\star(\alpha)$ be the solution of the the scaled problem (9). Then, we have:

$$x^\star(\alpha) = Q^{-1/2}\tilde{x}^\star(\alpha) = Q^{-1/2}\left(\left(\tilde{J}^\top\tilde{J} + \alpha I_n\right)^{-1}\tilde{J}^\top y\right) = Q^{-1/2}\underbrace{\left(\tilde{J}^\top\tilde{J} + \alpha I_n\right)^{-1}\tilde{J}^\top}_{\underset{\alpha\to 0}{\longrightarrow}\tilde{J}^\dagger} y$$

This proves that $x^\star(\alpha) \underset{\alpha\to 0}{\longrightarrow} Q^{-1/2}\tilde{J}^\dagger y =: x^\star$.

5. Prove that $x^\star$ is a solution of the non-regularized least squares problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|y - Jx\|_2^2 \tag{10}$$

Solution: Define the objective function $f(x) \coloneqq \frac{1}{2} \|y - Jx\|_2^2$. Then:

$$\begin{aligned}
\nabla f(x^\star) &= J^\top \left(y - Jx^\star\right), \\
&= J^\top \left(y - JQ^{-1/2}\tilde{J}^\dagger y\right), \\
&= Q^{1/2}\tilde{J}^\top \left(y - \tilde{J}\tilde{J}^\dagger y\right), \\
&= Q^{1/2}(\tilde{J}^\top - \underbrace{\tilde{J}^\top \tilde{J}\tilde{J}^\dagger}_{=\tilde{J}^\top})y, \\
&= Q^{1/2}\left(\tilde{J}^\top - \tilde{J}^\top\right)y, \\
&= 0
\end{aligned}$$

which proves that $x^\star$ is a stationary point of (10). Since it is a convex problem, this implies that $x^\star$ is a solution to (10).