Machine Learning Exercise Sheet 07

Constrained Optimization

Homework

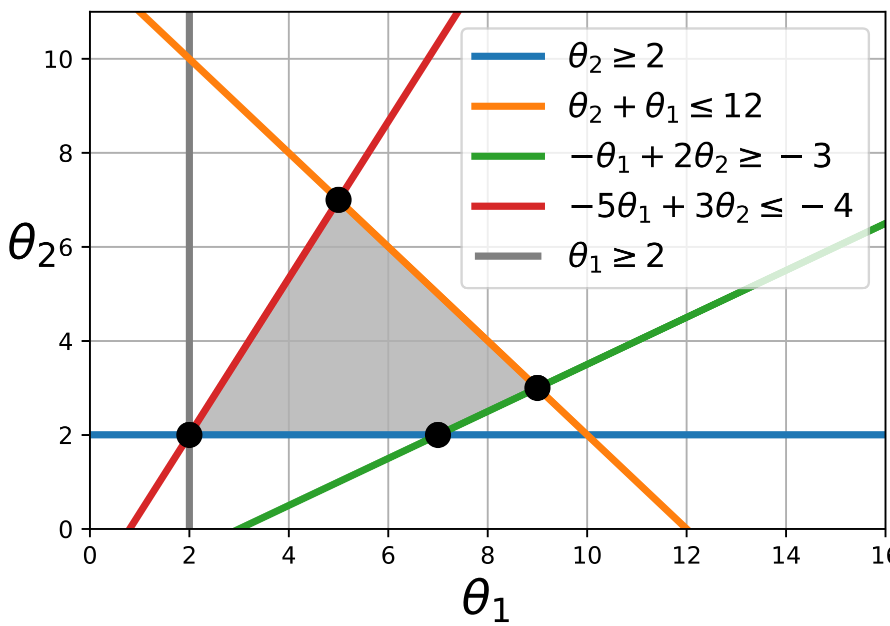
1 Constrained Optimization / Projected GD

Problem 1: Given is the following domain $\mathcal{X} \subset \mathbb{R}^2$ defined by a set of linear constraints

$$\begin{aligned}\mathcal{X} = \{\boldsymbol{\theta} \in \mathbb{R}^2 :\, & \theta_1 + \theta_2 \leq 12, \\ & -\theta_1 + 2\theta_2 \geq -3, \\ & -5\theta_1 + 3\theta_2 \leq -4, \\ & \theta_1 \geq 2,\ \theta_2 \geq 2\}\end{aligned}$$

and $f_0 : \mathbb{R}^2 \to \mathbb{R}$ with $f(\boldsymbol{\theta}) = 2\theta_1 - 3\theta_2$.

Visualize set $\mathcal{X}$ and find the minimizer $\boldsymbol{\theta}_{\min} \in \mathbb{R}^2$ and maximizer $\boldsymbol{\theta}_{\max} \in \mathbb{R}^2$ of f_0 on $\mathcal{X}$ as well as its minimum and maximum values. There is no need for a rigorous derivation in this exercise, but your plot should be nice and understandable and you still have to provide an explanation of your steps as well as clearly mention what results you apply.



The feasible set is a convex set and the objective function is linear, hence both convex and concave. We know from the lecture that the maximizer of a convex function and minimizer of a concave function lie on one of the vertices of the convex domain. That is, we only need to consider the points $(2,2)$, $(7,2)$, $(5,7)$ and $(9,3)$:

$$f_0(2,2) = -2, \quad f_0(7,2) = 8, \quad f_0(5,7) = -11, \quad f_0(9,3) = 9\,.$$

By evaluating f_0 at these 4 points we conclude that $\boldsymbol{\theta}_{\text{min}} = (5,7)$ and $\boldsymbol{\theta}_{\text{max}} = (9,3)$.

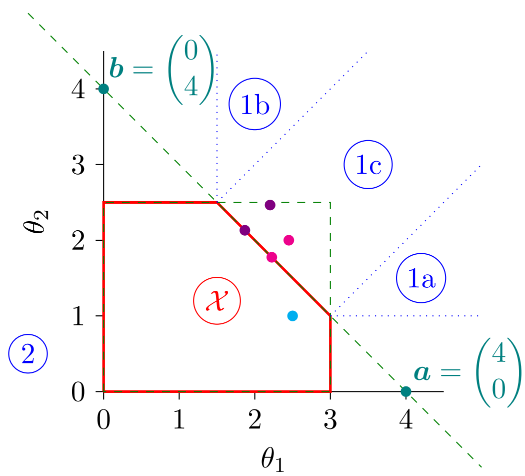
Problem 2: Given is the following domain $\mathcal{X} \subset \mathbb{R}^2$ defined by a set of linear constraints

$$\begin{aligned}\mathcal{X} = \{\boldsymbol{\theta} \in \mathbb{R}^2 :& \theta_1 + \theta_2 \le 4,\\ & 0 \le \theta_1 \le 3,\ 0 \le \theta_2 \le 2.5\}.\end{aligned}$$

a) Visualize set $\mathcal{X}$ and derive a closed form expression for the projection $\pi_{\mathcal{X}}(\boldsymbol{p}) = \arg\min_{\boldsymbol{\theta}\in\mathcal{X}} ||\boldsymbol{\theta}-\boldsymbol{p}||_2^2$. That is, given an arbitrary point $\boldsymbol{p} \in \mathbb{R}^2$, what is its projection on $\mathcal{X}$?

Hint: Use your plot from a) to divide $\mathbb{R}^2$ into regions. For one part of $\mathbb{R}^2$ you might want to use the following formula for projection on a hyperplane $\mathcal{X}_{a,b}$ defined by two points $\boldsymbol{a} \neq \boldsymbol{b} \in \mathcal{X}_{a,b}$ lying on it (note that in $\mathbb{R}^2$ two points uniquely determine a line):

$$\pi_{\mathcal{X}_{a,b}}(\boldsymbol{p}) = \boldsymbol{a} + \frac{(\boldsymbol{p}-\boldsymbol{a})^T(\boldsymbol{b}-\boldsymbol{a})}{\|\boldsymbol{b}-\boldsymbol{a}\|_2^2}(\boldsymbol{b}-\boldsymbol{a})$$



$$\pi_{\mathcal{X}}(\boldsymbol{p}) = \begin{cases} \boldsymbol{p} & \text{if } \boldsymbol{p} \in \mathcal{X} \\ (3,1)^T & \text{if } \boldsymbol{p} \in 1\text{a} = \{\boldsymbol{p} : p_2 \geq 1, \ -p_1 + p_2 \leq -2\} \\ (1.5, 2.5)^T & \text{if } \boldsymbol{p} \in 1\text{b} = \{\boldsymbol{p} : p_1 \geq 1.5, \ -p_1 + p_2 \geq 1\} \\ \pi_{\text{line}}(\boldsymbol{p}) & \text{if } \boldsymbol{p} \in 1\text{c} = \{\boldsymbol{p} : 2 < -p_1 + p_2 < 1, \ p_1 + p_2 > 4\} \\ \pi_{\text{box}}(\boldsymbol{p}) & \text{if } \boldsymbol{p} \in 2 = \{\boldsymbol{p} : p_1 < 1.5\} \cup \{\boldsymbol{p} : p_2 < 1\} \setminus \mathcal{X} \end{cases}$$

For π_{line} we use the provided hint with $\boldsymbol{a} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$.

$$\pi_{\text{line}}(\boldsymbol{p}) = \boldsymbol{a} + \frac{(\boldsymbol{p} - \boldsymbol{a})^T(\boldsymbol{b} - \boldsymbol{a})}{\|\boldsymbol{b} - \boldsymbol{a}\|_2^2}(\boldsymbol{b} - \boldsymbol{a}) = \begin{pmatrix} 4 \\ 0 \end{pmatrix} + \frac{1}{32}\left(\begin{pmatrix} p_1 - 4 \\ p_2 \end{pmatrix}^T \begin{pmatrix} -4 \\ 4 \end{pmatrix}\right)\begin{pmatrix} -4 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 4 \\ 0 \end{pmatrix} + \frac{16 - 4p_1 + 4p_2}{32}\begin{pmatrix} -4 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix} + \left(2 - \frac{1}{2}p_1 + \frac{1}{2}p_2\right)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{1}{2}p_1 - \frac{1}{2}p_2 \\ 2 - \frac{1}{2}p_1 + \frac{1}{2}p_2 \end{pmatrix}$$

$$\pi_{\text{box}}(\boldsymbol{p}) = \begin{pmatrix} \max(0, \min(3, p_1)) \\ \max(0, \min(2.5, p_2)) \end{pmatrix}$$

c) Perform two steps of projected gradient descent starting from the point $\boldsymbol{\theta}^{(0)} = (2.5, 1)^T$ with a constant learning rate of $\tau = 0.05$ for the following constrained optimization problem:

$$\begin{aligned} \text{minimize}_{\boldsymbol{\theta}} \quad & (\theta_1 - 2)^2 + (2\theta_2 - 7)^2 \\ \text{subject to} \quad & \boldsymbol{\theta} \in \mathcal{X} \end{aligned}$$

Projected gradient descent is a two-step algorithm consisting of regular gradient descent and projection:

1. $\boldsymbol{p}^{t+1} = \boldsymbol{\theta}^t - \tau \boldsymbol{\nabla}_{\boldsymbol{\theta}} f_0(\boldsymbol{\theta}^t)$
2. $\boldsymbol{\theta}^{t+1} = \pi_{\mathcal{X}}(\boldsymbol{p}^{t+1})$

The gradient of $f_0(\boldsymbol{\theta}) = (\theta_1 - 2)^2 + (2\theta_2 - 7)^2$ is

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} f_0(\boldsymbol{\theta}) = \begin{pmatrix} 2\theta_1 - 4 \\ 8\theta_2 - 28 \end{pmatrix}.$$

Using this gradient and the projection from above we can perform two steps of the algorithm:

$$\begin{aligned}
&1.1 \quad \boldsymbol{p}^{(1)} = \boldsymbol{\theta}^{(0)} - \tau \boldsymbol{\nabla}_{\boldsymbol{\theta}} f_0(\boldsymbol{\theta}^{(0)}) = \begin{pmatrix} 2.5 \\ 1 \end{pmatrix} - 0.05 \begin{pmatrix} 1 \\ -20 \end{pmatrix} = \begin{pmatrix} 2.45 \\ 2 \end{pmatrix} \\
&1.2 \quad \boldsymbol{\theta}^{(1)} = \pi_{\mathcal{X}}(\boldsymbol{p}^{(1)}) = \pi_{\text{line}}(\begin{pmatrix} 2.45 \\ 2 \end{pmatrix}) = \begin{pmatrix} 2.225 \\ 1.775 \end{pmatrix} \\
&2.1 \quad \boldsymbol{p}^{(2)} = \boldsymbol{\theta}^{(1)} - \tau \boldsymbol{\nabla}_{\boldsymbol{\theta}} f_0(\boldsymbol{\theta}^{(1)}) = \begin{pmatrix} 2.225 \\ 1.775 \end{pmatrix} - 0.05 \begin{pmatrix} 0.45 \\ -13.8 \end{pmatrix} = \begin{pmatrix} 2.2025 \\ 2.465 \end{pmatrix} \\
&2.2 \quad \boldsymbol{\theta}^{(2)} = \pi_{\mathcal{X}}(\boldsymbol{p}^{(2)}) = \pi_{\text{line}}(\begin{pmatrix} 2.2025 \\ 2.465 \end{pmatrix}) = \begin{pmatrix} 1.86875 \\ 2.13125 \end{pmatrix}
\end{aligned}$$

This process is also illustrated in the figure above, with points showing the starting point (cyan), step 1 (magenta) and step 2 (violet).

2 Lagrangian / Duality

Problem 3: Solve the following constrained optimization problem in $\mathbb{R}^2$ using the recipe described in the lecture (Slide 24).

$$\begin{aligned} \text{minimize}_{\boldsymbol{\theta}} \quad & \theta_1 - \sqrt{3}\theta_2 \\ \text{subject to} \quad & \theta_1^2 + \theta_2^2 - 4 \leq 0 \end{aligned}$$

Write down the **Lagrangian**

$$L(\boldsymbol{\theta}, \alpha) = \theta_1 - \sqrt{3}\theta_2 + \alpha\left(\theta_1^2 + \theta_2^2 - 4\right) .$$

For all $\alpha > 0$ function L is bounded from below and convex with respect to $\boldsymbol{\theta}$, so

$$g(\alpha) = \min_{\boldsymbol{\theta}} \theta_1 - \sqrt{3}\theta_2 + \alpha\left(\theta_1^2 + \theta_2^2 - 4\right)$$

can be computed setting the derivative (w.r.t. $\boldsymbol{\theta}$) to zero:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \alpha) = \begin{pmatrix} 1 + 2\alpha\theta_1 \\ -\sqrt{3} + 2\alpha\theta_2 \end{pmatrix} = 0 \quad \Rightarrow \quad \theta_1^*(\alpha) = -\frac{1}{2\alpha}, \quad \theta_2^*(\alpha) = \frac{\sqrt{3}}{2\alpha}.$$

Substituting this back into $L(\boldsymbol{\theta}^*(\alpha), \alpha)$ gives the dual function

$$g(\alpha) = -\frac{1}{2\alpha} - \frac{3}{2\alpha} + \frac{1}{4\alpha} + \frac{3}{4\alpha} - 4\alpha = -\frac{1}{\alpha} - 4\alpha \text{ if } \alpha > 0 \text{ and } g(0) = -\infty.$$

We must now maximize the dual function $g(\alpha)$ subject to $\alpha \geq 0$. Since $g(\alpha)$ is concave for $\alpha > 0$, we set its derivative to zero and solve for α.

$$\frac{dg(\alpha)}{d\alpha} = \frac{1}{\alpha^2} - 4 = 0 \quad \Rightarrow \quad \alpha^2 = \frac{1}{4} \quad \Rightarrow \quad \alpha^* = \frac{1}{2} \text{ since we require } \alpha \geq 0 .$$

The optimization problem is convex and Slater's condition holds (e.g. for $\boldsymbol{\theta} = \mathbf{0}$ we have $f_1(\mathbf{0}) = -4 < 0$), therefore the minimal value of f_0 is

$$p^* = d^* = g\left(\frac{1}{2}\right) = -4\,.$$

Optional: we can retrieve the minimizer $\boldsymbol{\theta}^*$ by substituting $\alpha^* = \frac{1}{2}$ into $\theta_1^*(\alpha) = -\frac{1}{2\alpha}$ and $\theta_2^*(\alpha) = \frac{\sqrt{3}}{2\alpha}$ and get

$$\theta_1^* = -1\,, \quad \theta_2^* = \sqrt{3}\,.$$

Problem 4: Given N data points $\boldsymbol{x}_i \in \mathbb{R}^d$ and their labels $y_i \in \{-1, 1\}$. Use results you know from the lecture and show that strong duality holds for the following problem.

$$\begin{aligned} \text{minimize}_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\ \text{subject to} \quad & y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) - 1 \geq 0 \quad \text{for } i = 1, \dots, N \end{aligned}$$

Note: by solving this problem we can find the separating hyperplane between points from different classes with the maximum margin, we will talk about it in the lecture when we cover SVMs.

The objective is convex and the constraints are affine in $\boldsymbol{w}$, thus the Slater's condition is satisfied, so the duality gap is zero.

In-class Exercises

3 Duality

Problem 5: Given N numbers $x_1, ..., x_N \in \mathbb{R}$, construct a minimization problem such that its optimal value is $\max(x_1, \ldots, x_N)$ and derive the Lagrange dual problem.

The following equation finds the maximum via optimization:

$$\begin{aligned} &\text{minimize}_b \quad b \\ &\text{subject to} \quad b \geq x_i \;\Leftrightarrow\; x_i - b \leq 0, \quad i = 1, \ldots, N\,. \end{aligned}$$

Intuitively, we define an upper limit b and reduce it as much as possible while being greater or equal than all values in our set. We can now apply our recipe for solving a constrained optimization problem, with parameters $\boldsymbol{\theta} = b$ and Lagrange multipliers $\boldsymbol{\alpha} = \boldsymbol{w}$:

1. Calculate the Lagrangian:

$$L(b, \boldsymbol{w}) = b + \sum_{i=1}^{N} w_i (x_i - b) = b\left(1 - \sum_{i=1}^{N} w_i\right) + \sum_{i=1}^{N} w_i x_i$$

2. Obtain the Lagrange dual function g by solving the unconstrained problem $\min_b L(b, \boldsymbol{w})$. Note that $L(b, \boldsymbol{w})$ becomes unbounded below with respect to $b \in \mathbb{R}$ if $1 - \sum_{i=1}^{N} w_i \neq 0$ (in this case we have a non-constant affine function) and does not depend on b otherwise since b is being multiplied by 0. Therefore,

$$g(\boldsymbol{w}) = \begin{cases} \sum_{i=1}^{N} w_i x_i & \text{if } 1 - \sum_{i=1}^{N} w_i = 0\,, \\ -\infty & \text{otherwise}\,. \end{cases}$$

3. The dual problem is then

$$\begin{aligned} \text{maximize}_{\boldsymbol{w}} \quad & \sum_{i=1}^{N} w_i x_i \\ \text{subject to} \quad & \sum_{i=1}^{N} w_i = 1 \\ & w_i \geq 0 \quad \text{for } i = 1, \ldots, N\,. \end{aligned}$$

The dual problem we've found is another intuitive solution of our problem. We assign weights w_i to all elements in our set. By maximizing the sum $\sum_{i=1}^{N} w_i x_i$ while keeping the sum of all w_i's constant, the full weight will be assigned to the largest element. The solution w_i^* will then be 1 for the index i of the maximum element end 0 otherwise.

Problem 6: Given N numbers $x_1, ..., x_N \in \mathbb{R}$, construct a maximization problem such that its optimal value is the sum of the k largest values and derive the Lagrange dual problem.

Defining a minimization problem as we did above is not as obvious in this case. So, instead we start with a maximization problem similar to the one we derived before:

$$\begin{aligned}
\text{maximize}_{\boldsymbol{w}} \quad & \sum_{i=1}^{N} w_i x_i \\
\text{subject to} \quad & \sum_{i=1}^{N} w_i = k \\
& w_i - 1 \leq 0 \quad \text{for } i = 1, \ldots, N \\
& -w_i \leq 0 \quad \text{for } i = 1, \ldots, N\,.
\end{aligned}$$

Since all weights w_i are constrained to be between 0 and 1, the solution of this problem will find the indices of the largest elements and return their sum. This time, we will apply our recipe in the opposite direction to find the Lagrange dual problem, with parameters $\boldsymbol{\theta} = \boldsymbol{w}$ and Lagrange multipliers $\boldsymbol{\alpha} = (b, \boldsymbol{s}, \boldsymbol{t})$ (note that the dual multipliers b corresponding to equality constraints are *unconstrained* in the dual problem):

1. Calculate the Lagrangian:

$$\begin{aligned}
L(\boldsymbol{w}, b, \boldsymbol{s}, \boldsymbol{t}) &= -\sum_{i=1}^{N} w_i x_i + b\left(\sum_{i=1}^{N} w_i - k\right) + \sum_{i=1}^{N} s_i(w_i - 1) - \sum_{i=1}^{N} t_i w_i \\
&= \sum_{i=1}^{N} w_i(-x_i + b + s_i - t_i) - bk - \sum_{i=1}^{N} s_i
\end{aligned}$$

   Note that since we are maximizing we need to flip the sign of the function (since the maximizer of f is the minimizer of $-f$).

2. Obtain the Lagrange dual function $g(b, \boldsymbol{s}, \boldsymbol{t}) = \min_{\boldsymbol{w}} L(\boldsymbol{w}, b, \boldsymbol{s}, \boldsymbol{t})$ similarly as in the previous exercise:

$$g(b, \boldsymbol{s}, \boldsymbol{t}) = \begin{cases} -bk - \sum_{i=1}^{N} s_i & \text{if } -x_i + b + s_i - t_i = 0 \text{ for all } i = 1, \ldots, N\,, \\ -\infty & \text{otherwise}\,. \end{cases}$$

3. Since all t_i's are non-negative numbers and not important for the final solution, we can write the above condition as $-x_i + b + s_i = t_i \geq 0$. Using this, we can write the dual problem as

$$\begin{aligned}
\text{maximize}_{b,\boldsymbol{s},\boldsymbol{t}} \quad & -kb - \sum_{i=1}^{N} s_i \\
\text{subject to} \quad & b \geq x_i - s_i \quad \text{for } i = 1, \ldots, N \\
& s_i \geq 0 \quad \text{for } i = 1, \ldots, N\,.
\end{aligned}$$

Note that we don't have any constraints on the Lagrange multiplier b since it stems from an equality constraint. We can now simply flip the sign to obtain a minimization problem:

$$\begin{aligned} \text{minimize}_{b,\boldsymbol{s},\boldsymbol{t}} \quad & kb + \sum_{i=1}^{N} s_i \\ \text{subject to} \quad & b \geq x_i - s_i \quad \text{for } i = 1, \dots, N \\ & s_i \geq 0 \quad \text{for } i = 1, \dots, N \,. \end{aligned}$$

We have now found a problem formulation that is analogous to the one we used in the previous problem! Intuitively, b returns the value of the k'th largest element. The values s_i return the difference of the $k-1$ larger elements to this one and are 0 for all smaller or equal elements. Hence, the above sum is the sum of the k largest elements.

Problem 7: Given $\boldsymbol{c} \in \mathbb{R}^d$, $\boldsymbol{b} \in \mathbb{R}^M$ and $\boldsymbol{A} \in \mathbb{R}^{M \times d}$, derive the Lagrange dual problem of

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \,. \end{aligned}$$

- Lagrangian: $L(\boldsymbol{x}, \boldsymbol{\alpha}) = \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{\alpha}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{x}^T (\boldsymbol{A}^T \boldsymbol{\alpha} + \boldsymbol{c}) - \boldsymbol{b}^T \boldsymbol{\alpha}$ is a linear non-constant function if $\boldsymbol{A}^T \boldsymbol{\alpha} + \boldsymbol{c} \neq 0$ (and therefore unbounded below with respect to x in this case) and does not depend on x otherwise.
- Dual function:

$$g(\boldsymbol{\alpha}) = \min_{\boldsymbol{x} \in \mathbb{R}^d} \boldsymbol{x}^T (\boldsymbol{A}^T \boldsymbol{\alpha} + \boldsymbol{c}) - \boldsymbol{b}^T \boldsymbol{\alpha} = \begin{cases} -\boldsymbol{b}^T \boldsymbol{\alpha} & \text{if } \boldsymbol{A}^T \boldsymbol{\alpha} + \boldsymbol{c} = \boldsymbol{0} \\ -\infty & \text{otherwise} \,. \end{cases}$$

- Dual problem:

$$\begin{aligned} \text{maximize}_{\boldsymbol{\alpha}} \quad & -\boldsymbol{b}^T \boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{A}^T \boldsymbol{\alpha} + \boldsymbol{c} = \boldsymbol{0} \\ & \boldsymbol{\alpha} \geq \boldsymbol{0} \end{aligned}$$

Problem 8: Given $\boldsymbol{b} \in \mathbb{R}^M$, $\boldsymbol{A} \in \mathbb{R}^{M \times d}$ and positive definite $\boldsymbol{Q} \in \mathbb{R}^{d \times d}$, derive the Lagrange dual problem of

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & \frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \,. \end{aligned}$$

- Lagrangian: $L(\boldsymbol{x}, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\alpha}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$, note that $L(\boldsymbol{x}, \boldsymbol{\alpha})$ is strictly convex (hence bounded below) with respect to $\boldsymbol{x}$ for all values of $\boldsymbol{\alpha}$ thanks to the quadratic part.
- Dual function: $g(\boldsymbol{\alpha}) = \min_{\boldsymbol{x}\in\mathbb{R}^d} \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\alpha}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$ can be found by solving a non-constrained strictly convex optimization problem (compute the gradient, set it to zero).

$$\boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\alpha}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{a} = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{x}^*(\boldsymbol{\alpha}) = -\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\alpha}$$

  And finally $g(\boldsymbol{\alpha}) = L(\boldsymbol{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = -\frac{1}{2}\boldsymbol{\alpha}^T\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\right)\boldsymbol{\alpha} - \boldsymbol{b}^T\boldsymbol{\alpha}$.
- Dual problem:

$$\begin{aligned} \text{maximize}_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\right)\boldsymbol{\alpha} - \boldsymbol{b}^T\boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{\alpha} \geq \boldsymbol{0} \end{aligned}$$

  or equivalently

$$\begin{aligned} \text{minimize}_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}^T\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\right)\boldsymbol{\alpha} + \boldsymbol{b}^T\boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{\alpha} \geq \boldsymbol{0} \end{aligned}$$

First, note the primal problem was convex and Slater's constraint qualification hold assuring the strong duality. The dual task is now again a convex QP (one can show that the matrix $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T$ is again positive definite), but with M variables and M simple non-negativity constraints. It comes in cost of having the inverse $\boldsymbol{Q}^{-1}$ in the dual formulation, so one should consider solving the dual instead of the primal especially if $M << d$ (e.g. only one constraint in the initial task) and when $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T$ can be computed efficiently based on some additional knowledge about $\boldsymbol{Q}$ and $\boldsymbol{A}$.

4 Constrained Optimization: Toy Problem

Problem 9: Suppose we have 40 pieces of raw material. Toy A can be made of one piece material with 3 EUR machining fee. A larger toy B can be made from two pieces of material with 5 EUR machining fee.

Because distribution costs decrease with larger quantities, we can sell x pieces of toy A for $20 - x$ EUR each, and y pieces of toy B for $40 - y$ EUR each. From our experience, toy B is more popular than toy A; therefore, we will produce not more of toy A than of toy B.

To get the maximum profit, we want to calculate the amount of toy A and toy B that we should produce.

a) Write down the constrained optimization problem and the associated Lagrangian.

b) Write down the Karush–Kuhn–Tucker (KKT) conditions for the above optimization problem.

c) Obtain the solution to the constrained optimization problem by solving the KKT conditions. Do not worry about non-integer production quantities.

a) The constrained optimization problem is

$$\begin{aligned} \min f(x,y) &= -[x(20-x)+y(40-y)-3x-5y] = x^2-17x+y^2-35y \\ s.t.\ f_1(x,y) &= x+2y-40 \le 0 \\ f_2(x,y) &= x-y \le 0 \end{aligned}$$

and the associated Lagrangian is given by

$$L(x,y,\alpha_1,\alpha_2) = x^2-17x+y^2-35y+\alpha_1(x+2y-40)+\alpha_2(x-y)$$

with $\alpha_1 \ge 0$ and $\alpha_2 \ge 0$.

b) Primal feasibility:

$$\begin{aligned} x+2y-40 &\le 0 \\ x-y &\le 0 \end{aligned}$$

Dual feasibility:

$$\begin{aligned} \alpha_1 &\ge 0 \\ \alpha_2 &\ge 0 \end{aligned}$$

Complementary slackness:

$$\begin{aligned} \alpha_1(x+2y-40) &= 0 \\ \alpha_2(x-y) &= 0 \end{aligned}$$

x, y minimize Lagrangian:

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2x-17+\alpha_1+\alpha_2 = 0 \\ \frac{\partial L}{\partial y} &= 2y-35+2\alpha_1-\alpha_2 = 0 \end{aligned}$$

c) We start by observing that the KKT complementary slackness conditions demand that either $\alpha_1 = 0$ or $x+2y=40$ (i.e. production is limited by available resources) and $\alpha_2 = 0$ or $x = y$ (i.e. production is limited by our desire to not produce more of toy A than toy B). We *guess* that the production will be limited by available resources and not by our desire to produce more of toy A then toy B, thus

$$\begin{aligned} x+2y-40 &= 0 \\ \alpha_2 &= 0\,. \end{aligned}$$

Solving the first condition for x gives $x = 40-2y$.

Substituting this expression for x together with $\alpha_2 = 0$ into the KKT minimization conditions ($\partial L/\partial x = 0$, $\partial L/\partial y = 0$) and solving for y and then x gives

$$\begin{aligned} y &= 16.1 \\ x &= 7.8\,. \end{aligned}$$

We now have to check the correctness of our assumptions by verifying the remaining KKT conditions. Since we explicitly used $x + 2y = 40$, the first primal feasibility condition is satisfied with equality. We explicitly check that $x \leq y$ is satisfied. We assumed $\alpha_2 = 0$, thus the second dual feasibility condition is satisfied. To check that $\alpha_1 \geq 0$, we solve the KKT minimization conditions for α_1, insert our values for x, y and obtain $\alpha_1 = 1.4 > 0$.

Consequently our guess was correct and we obtained a solution for the constrained optimization problem from the KKT conditions. If any constraint would have been violated, we would have to try a new combination of active/inactive constraints. Since there are 2^N combinations of active/inactive constraints for N constraints, the method shown here is only effective for reasonably small N.