V. Schulz
M. Klar

SS 2022
Submission: 03.05.21, until 12:15

Numerical Optimization - Sheet 4

If you are a student in mathematics please solve the exercises with no tag and the ones with the tag **Mathematics**. If you are a data science student please solve the problems with no tag and those with the tag **Data Science**. Submissions with tags other than your subject count as bonus points. The tag **Programming** marks programming exercises.

Ex 1 (2 Points)

Let $x \in \mathbb{R}^2$. Use the necessary and sufficient conditions for optimality to solve

$$\begin{aligned} &\min_x x_1 + x_2 \\ &\text{s.t. } x_1^2 + x_2^2 = 2. \end{aligned}$$

Solution 1:
Due to $\mathcal{L} = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2)$ and $\nabla_x \mathcal{L} = (1 + \lambda 2x_1, 1 + \lambda 2x_2)$ we obtain from the KKT-conditions

$$\begin{aligned} 1 + \lambda 2x_1 &= 0 \\ 1 + \lambda 2x_2 &= 0. \end{aligned}$$

The constraints yield

$$x_1^2 + x_2^2 = 2.$$

Therefore

$$x_1 = x_2 = -\frac{1}{2\lambda}$$

and an application of the constrains imply

$$\lambda = \pm\frac{1}{2}.$$

As

$$\nabla^2 \mathcal{L} = \begin{bmatrix} 2\lambda & 0 \\ 0 & 2\lambda \end{bmatrix}$$

is positive definite only if $\lambda = \frac{1}{2}$, we obtain the solution

$$\hat{x} = (-1, -1).$$

Ex 2 (4 Points)

Find a local minimum of the problems.

(i) Let $c > 0$ be fixed.

$$\begin{aligned} \min_{r,h\in\mathbb{R}} \; & r^2 + rh \\ \text{s.t. } & \pi r^2 h = c \\ & r \geq 0. \end{aligned}$$

(ii)

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \; & \frac{1}{2} x^\top \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} x - x^\top \begin{pmatrix} 14 \\ 21 \end{pmatrix} \\ \text{s.t. } & \begin{bmatrix} 1, -2 \end{bmatrix} x = 0. \end{aligned}$$

<u>Solution 2:</u>

(i) The constraints yield that $r \neq 0$. We can therefore substitute

$$\frac{c}{\pi r^2} = h.$$

So we rewrite the problem

$$\begin{aligned} \min_{r,h\in\mathbb{R}} \; & r^2 + rh \\ \text{s.t. } & \pi r^2 h = c \\ & r \geq 0. \end{aligned}$$

as

$$\begin{aligned} \min_{r\in\mathbb{R}} \; & r^2 + \frac{c}{\pi r} =: f(r) \\ \text{s.t. } & r \geq 0. \end{aligned}$$

We now search a stationary point r^* of f and check whether it fulfills the constraints $r^* \geq 0$ in a second step. The necessary condition

$$f'(r^*) = 2r^* - \frac{c}{\pi {r^*}^2} = 0,$$

gives

$$r^* = \left(\frac{c}{\pi 2}\right)^{\frac{1}{3}} \geq 0,$$

which also fulfills $r^* \geq 0$. The constraints in the original problem finally show that

$$h^* = \left(\frac{2c^2}{\pi^2}\right)^{\frac{1}{3}}.$$

(ii)

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^\top \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} x - x^\top \begin{pmatrix} 14 \\ 21 \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ -2 \end{pmatrix}^\top x$$

1. Order Conditions
We obtain the KKT Conditions

$$4x_1 + 6x_2 - 14 - \lambda = 0 \tag{1}$$
$$6x_1 + 9x_2 - 21 + \lambda 2 = 0 \tag{2}$$

If we subtract $\frac{2}{3}(1)$ from (2) we obtain $\lambda = 0$. The rows in the Hessian of the objective are linearly dependent.

Constraints
The constraints yield $x_1 = 2x_2$. Therefore (1) yields

$$8x_2 + 6x_2 = 14 \Rightarrow x_2 = 1$$

Which implies $x_1 = 2$.

2. Order Sufficient Cond. We now fix the vector $\tilde{x} = (1, 2)$ of which we know that $\text{span}(\tilde{x}) = \ker(C)$ as $\ker(C)$ is 1-dimensional and we find that

$$\tilde{x}^\top \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} \tilde{x} = 64 > 0.$$

So the 2. derivative of the objective is positive definite on the kernel of the constraints and the above solution is a strict local minimum.

Ex 3 (6 Points)

Let $C > 0$. Compute the Wolfe-Dual of the optimization problem

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^m \xi_i$$
s.t.
$$y_i(w^T x_i + b) \geq 1 - \xi_i,\ \forall i = 1, \ldots, m$$
$$\xi_i \geq 0,\ \forall i = 1, \ldots, m.$$

Solution 3:
The Lagrangian is given by

$$\mathcal{L}(w, b, \xi, \beta, \gamma) = \frac{1}{2}w^T w + C\sum_{i=1}^m \xi_i + \sum_{i=1}^m \beta_i(1 - \xi_i - y_i(w^T x_i + b)) - \sum_{i=1}^m \gamma_i \xi_i.$$

The constraints for the Wolfe dual are

$$\frac{\partial \mathcal{L}}{\partial w}(w) = w^T - \sum_{i=1}^m \beta_i y_i x_i^T = 0 \tag{3}$$
$$\frac{\partial \mathcal{L}}{\partial b}(b) = -\sum_{i=1}^m \beta_i y_i = 0 \tag{4}$$
$$\frac{\partial \mathcal{L}}{\partial \xi} = \mathbf{1}^T C - \beta^T - \gamma^T = 0 \tag{5}$$
$$\beta \geq 0,\ \gamma \geq 0. \tag{6}$$

The Wolfe dual is defined as

$$\max_{w,b,\xi,\beta,\gamma} \mathcal{L}(w, b, \xi, \beta, \gamma)$$
$$\text{s.t. } \nabla \mathcal{L}_{w,b,\xi} = 0$$
$$\beta, \gamma \geq 0.$$

Substituting w and elimination of b, ξ and γ yield

$$
\begin{aligned}
\mathcal{L}(w,b,\xi,\beta,\gamma) &= \frac{1}{2}w^T w + C\sum_{i=1}^m \xi_i + \sum_{i=1}^m \beta_i(1-\xi_i - y_i(w^T x_i + b)) - \sum_{i=1}^m \gamma_i \xi_i \qquad \text{(s. 1)}\\
&= \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + C\sum_{i=1}^m \xi_i + \sum_{i=1}^m \beta_i\left(1-\xi_i - y_i\left(\sum_{j=1}^m \beta_j y_j x_j^T x_i + b\right)\right) - \sum_{i=1}^m \gamma_i\xi_i =\\
&= \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + C\mathbf{1}^T\xi + \sum_{i=1}^m \beta_i - \beta^T\xi - \sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j - \sum_{i=1}^m y_i\beta_i b - \gamma^T\xi =\\
&= \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + C\mathbf{1}^T\xi - \beta^T\xi - \gamma^T\xi + \sum_{i=1}^m \beta_i - \sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j - \sum_{i=1}^m y_i\beta_i b = \qquad \text{(s. 2)}\\
&= \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + C\mathbf{1}^T\xi - \beta^T\xi - \gamma^T\xi + \sum_{i=1}^m \beta_i - \sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j = \qquad \text{(s. 3)}\\
&= \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + \sum_{i=1}^m \beta_i - \sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j =\\
&= -\frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \beta_i\beta_j y_i y_j x_i^T x_j + \sum_{i=1}^m \beta_i.
\end{aligned}
$$

Furthermore we obtain from (5) and (6) that

$$
\begin{aligned}
\mathbf{1}C - \beta = \gamma \geq 0\\
\Rightarrow 0 \leq \beta \leq C.
\end{aligned}
$$

We are left with the variables β und y for which we copy the remaining constraints. Finally the Wolfe dual reduces to

$$
\begin{aligned}
&\max_\beta \sum_{i=1}^m \beta_i - \frac{1}{2}\sum_{i,j=1}^m \beta_i\beta_j x_i^T x_j y_i y_j\\
&\text{s.t} \sum_{i=1}^m \beta_i y_i = 0\\
&0 \leq \beta \leq C.
\end{aligned}
$$

Ex 4 Programming (3 Points)

You find the iPython notebook `numopt_version02` in the folder Lecture 6 on Olat. Use this notebook as basis for the following exercise.

(i) Consider the cell **chapter 2 SVM example code**. Uncomment the line which generates the *"not nicely separable points"* and fit a *linear* SVM using a hard and a soft margin. Fit a SVM with *rbf*-Kernel, again with hard and soft margin. Visualize your results with the code given in the notebook.

(ii) The cell **chapter 3 3D pic for SVM illustration** visualizes the feature map

$$(x_1, x_2) \mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2).$$

The function `svm.SVC` allows user defined kernels. Determine and implement the corresponding kernel function and fit and visualize the model to the data of (i).

(iii) Please comment on the result and the capabilites of the given kernel.