Question 1 [20 points] Given a function $f$ of two vector variables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, suppose that $(x^*, y^*)$ is a saddle point:

$$f(x^*, y^*) = \max_{y \in \mathcal{Y}} f(x^*, y) = \min_{x \in \mathcal{X}} f(x, y^*).$$

Without assuming any other properties of $f$, show that $f$ satisfies the minimax property

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x, y).$$

Solution.
For any $a \in \mathcal{X}$ and $b \in \mathcal{Y}$ we have

$$\min_{y \in \mathcal{Y}} f(a, y) \leq f(a, b) \leq \max_{x \in \mathcal{X}} f(x, b).$$

It follows that

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y) \leq f(a, b) \leq \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y).$$

We use the existence of a saddle point to prove the converse inequality. To this end we have:

$$\min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y) \leq \max_{x \in \mathcal{X}} f(x, y^*) = \min_{y \in \mathcal{Y}} f(x^*, y) \leq \max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y).$$

Question 2 [30 points] Consider the optimization problem for SVMs with soft margins. Suppose we replace the penalty term $\sum_{i=1}^{m} \xi_i$ with $\max_{i=1}^{m} \xi_i$. Give the associated dual problem. Show that it differs from the standard dual optimization problem only by the constraints.

Solution. The optimization problem is as follows:

$$\min_{\vec{w}, b, \vec{\mu}} \frac{1}{2} ||\vec{w}||^2 + C\xi$$

s.t. $y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi$, for $i = 1, \ldots, m$

$\xi \geq 0$

where $C$ is a parameter of the algorithm.

To dualise this problem we consider first the Lagrangian. This is defined by

$$L(\vec{w}, b, \xi, \vec{\alpha}, \beta) = \frac{1}{2} ||\vec{w}||^2 + C\xi - \sum_{i=1}^{m} \alpha_i [y_i(\vec{w} \cdot \vec{x}_i + b) - 1 + \xi] - \beta \xi,$$
where $\vec{\alpha}$ and $\beta$ are the Lagrange multipliers.

To obtain an expression for

\[
\min_{\vec{w}, b, \vec{\xi}} L(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \beta)
\]

we set the gradient of the Lagrangian with respect to the primal variables $\vec{w}, b, \vec{\xi}$ to zero. This yields

\[
\nabla_{\vec{w}} L = \vec{w} - \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i = 0 \implies \vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i
\]

(1)

\[
\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i = 0 \implies \sum_{i=1}^{m} \alpha_i y_i = 0
\]

(2)

\[
\frac{\partial L}{\partial \vec{\xi}} = C - \sum_{i=1}^{m} \alpha_i - \beta = 0 \implies \sum_{i=1}^{m} \alpha_i = C - \beta
\]

(3)

Substituting $\vec{w} = \sum_{i=1}^{m} \alpha_i y_i \vec{x}_i$ into the Lagrangian and using equations (2) and (3) to simplify the resulting expression by eliminating the terms in $\beta$ and $\vec{\xi}$, we obtain the dual problem:

\[
\max \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j \vec{x}_i \cdot \vec{x}_j
\]

s.t. $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$.

Note that the equation $\sum_{i=1}^{m} \alpha_i = C - \beta$ and constraint $\beta \geq 0$ were translated to the inequality $\sum_{i=1}^{m} \alpha_i \leq C$ in the above problem.

**Question 3 [50 points]**

The Sequential Minimal Optimization (SMO) algorithm is a step-by-step method to reduce large quadratic programming optimization problems to a series of small optimizations involving only two Lagrange multipliers. In this question we derive the update rule for the SMO algorithm in the context of the dual formulation of the soft-margin SVM problem:

\[
\max_{\vec{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)
\]

\[
\Psi_{\alpha_1, \alpha_2}(\vec{\alpha})
\]

subject to: $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$, $i \in \{1, \ldots, m\}$. 

2
(a) Assume that we want to carry out the optimization above only over $\alpha_1$ and $\alpha_2$, keeping all other variables fixed. Show that the optimization reduces to
\[
\max_{\alpha_1, \alpha_2} \alpha_1 + \alpha_2 - \frac{1}{2} K_{11} \alpha_1^2 - \frac{1}{2} K_{22} \alpha_2^2 - s K_{12} \alpha_1 \alpha_2 - y_1 \alpha_1 v_1 - y_2 \alpha_2 v_2
\]
subject to: $0 \leq \alpha_1, \alpha_2 \leq C \land \alpha_1 + s \alpha_2 = \gamma$,
where $s = y_1 y_2$, $\gamma = y_1 \sum_{i=3}^m y_i \alpha_i$, $K_{ij} = x_i \cdot x_j$, and $v_i = \sum_{j=1}^m \alpha_j y_j K_{ij}$ for $i = 1, 2$.

(b) Substitute the linear constraint $\alpha_1 = \gamma - s \alpha_2$ into $\Psi$ above to obtain a new objective function $\Psi_2$ that depends only on $\alpha_2$. Show that the $\alpha_2$ that minimises $\Psi_2$ (ignoring the constraints $0 \leq \alpha_1, \alpha_2 \leq C$) can be expressed as
\[
\alpha_2 = \frac{s(K_{11} - K_{12}) \gamma + y_2 (v_1 - v_2) - s + 1}{\eta},
\]
where $\eta = K_{11} + K_{22} - 2K_{12}$.

(c) Writing $\alpha_i^*$ for the values of the $\alpha_i$ prior to optimization over $\alpha_1, \alpha_2$, show that
\[
v_1 - v_2 = f(x_1) - f(x_2) + \alpha_2^* y_2 \eta - y_1 \gamma (K_{11} - K_{12})
\]
where $f(x) = \sum_{i=1}^m \alpha_i^* y_i (x_i \cdot x)$.

(d) Show that
\[
\alpha_2 = \alpha_2^* + \frac{y_2 (f(x_2) - f(x_1))}{H} (y_2 - y_1) .
\]

(e) If $s = 1$, define $L = \max\{0, \gamma - C\}$ and $H = \min\{C, \gamma\}$ as the lower and upper bounds on $\alpha_2$. The update rule for SMO involves “clipping” the value of $\alpha_2$, i.e.,
\[
\alpha_2^{\text{clip}} = \begin{cases} 
\alpha_2 & \text{if } L < \alpha_2 < H \\
L & \text{if } \alpha_2 \leq L \\
H & \text{if } \alpha_2 \geq H.
\end{cases}
\]
We subsequently solve for $\alpha_1$ in the equality constraint $\alpha_1 = \alpha_1^* + s(\alpha_2^* - \alpha_2^{\text{clip}})$. Why is clipping required? How do we define $L$ and $H$ in the case $s = -1$?

**Solution.**

1. The new objective function is obtained by removing all terms that don’t depend on $\alpha_1$ and $\alpha_2$ from the objective function in the original problem.
2. Substituting into $\Psi_1$, we have:
\[
\Psi_2 = \gamma - s \alpha_2 + \alpha_2 - \frac{1}{2} K_{11} (\gamma - s \alpha_2)^2 - \frac{1}{2} K_{22} \alpha_2^2 - s K_{12} (\gamma - s \alpha_2) \alpha_2 - y_1 (\gamma - s \alpha_2) v_2 - y_2 \alpha_2 v_2 .
\]
At the stationary point we have
\[
\frac{d\Psi_2}{d\alpha_2} = -s + 1 + s K_{11} (\gamma - s \alpha_2) - K_{22} \alpha_2 - s K_{12} (\gamma - s \alpha_2) + s^2 K_{12} \alpha_2 + y_1 s v_1 - y_2 v_2 = 0 .
\]
Noting that $s^2 = 1$ and rearranging terms we get the desired equation for $\alpha_2$.  

3. By definition of $f$, we see that $v_1 = f(x_1) - y_1\alpha_1^* K_{11} - y_2\alpha_2^* K_{12}$ and similarly $v_2 = f(x_2) - y_1\alpha_1^* K_{11} - y_2\alpha_2^* K_{22}$. Using these equations together with the identities $\alpha_1^* = \gamma - s\alpha_2^*$ and $y_1 = sy_2$ we have

$$v_1 - v_2 = f(x_1) - f(x_2) - y_1\alpha_1^*(K_{11} - K_{12}) - y_2\alpha_2^*(K_{12} - K_{22})$$

$$= f(x_1) - f(x_2) - y_1(\gamma - s\alpha_2^*) (K_{11} - K_{12}) - y_2\alpha_2^*(K_{12} - K_{22})$$

$$= f(x_1) - f(x_2) - sy_2\gamma (K_{11} - K_{12}) - y_2\alpha_2^*(K_{11} - K_{12}) - y_2\alpha_2^*(K_{12} - K_{22})$$

$$= f(x_1) - f(x_2) - sy_2\gamma (K_{11} - K_{12}) + y_2\alpha_2^*$$. 

4. Combining the results from (b) and (c), we have

$$\eta\alpha_2 = s(K_{11} - K_{12})\gamma + y_2 [f(x_1) - f(x_2) - sy_2\gamma (K_{11} - K_{12}) + y_2\alpha_2^*\eta] - s + 1$$

$$= y_2 [f(x_1) - f(x_2) + y_2\alpha_2^*\eta] - s + 1$$

$$= \alpha_2^*\eta + y_2 [f(x_1) - f(x_2) - y_1 + y_2]$$

$$= \alpha_2^*\eta + y_2 [(y_2 - f(x_2)) - (y_1 - f(x_1))]$$.

Dividing both sides by $\eta$ yields the desired result.

5. Clipping ensures that the new values of $\alpha_1$ and $\alpha_2$ satisfy the inequality constraints $0 \leq \alpha_1, \alpha_2 \leq C$.

If $s = -1$ then we define $L = \max(0, -\gamma)$ and $H = \min(C, C - \gamma)$. 

4