Lagrange multipliers and optimization problems

We'll present here a very simple tutorial example of using and understanding Lagrange multipliers. Let w be a scalar parameter we wish to estimate and x a fixed scalar. We wish to solve the following (tiny) SVM like optimization problem:

$$\text{minimize } \frac{1}{2}w^2 \text{ subject to } wx - 1 \geq 0 \tag{1}$$

This is difficult only because of the constraint. We'd rather solve an unconstrained version of the problem but, somehow, we have to take into account the constraint. We can do this by including the constraint itself in the minimization objective as it allows us to twist the solution towards satisfying the constraint. We need to know how much to emphasize the constraint and this is what the Lagrange multiplier is doing. We will denote the Lagrange multiplier by α to be consistent with the SVM problem. So we have now constructed a new minimization problem (still minimizing with respect to w) that includes the constraint as an additional linear term:

$$J(w;\alpha) = \frac{1}{2}w^2 - \alpha(wx - 1) \tag{2}$$

The Lagrange multiplier α appears here as a parameter. You might view this new objective a bit suspiciously since we appear to have lost the information about what type of constraint we had, i.e., whether the constraint was $wx - 1 \geq 0$, $wx - 1 \leq 0$, or $wx - 1 = 0$. How is this information encoded? We can encode this by constraining the values of the Lagrange multipliers:

$$\begin{array}{lcl} wx - 1 \geq 0 & \Rightarrow & \alpha \geq 0 \\ wx - 1 \leq 0 & \Rightarrow & \alpha \leq 0 \\ wx - 1 = 0 & \Rightarrow & \alpha \text{ is unconstrained} \end{array}$$

Note, for example, that when the constraint is $wx - 1 \geq 0$, as we have above, large positive values of α will encourage choices of w that result in large positive values for $wx - 1$. This is because in the above objective, $J(w;\alpha)$, we try to minimize $-\alpha(wx - 1)$ in addition to $w^2/2$; minimizing $-\alpha(wx - 1)$ is the same as maximizing $\alpha(wx - 1)$ or $wx - 1$ since α is positive. Figure 1 tries to illustrate this effect. Assuming $x = 1$ we can plot the new objective function as a function of w for different values of α. Larger values of α clearly move the solution (minimizing w) towards satisfying $w - 1 \geq 0$ (we assume here that $x = 1$). Based on the figure we can see that setting $\alpha = 1$ produces just the right solution, i.e., $w^* = 1$, which satisfies the constraint $wx - 1 \geq 0$ (when $x = 1$) with minimal distortion of the original objective. There's no reason to consider negative values for α since they would push the solution away from satisfying our inequality constraint.

Effectively what we are doing here is solving a large number of optimization problems, once for each setting of the Lagrange multiplier α. Indeed, we can express the solution (the minimizing w) as a parametric function of α:

$$\frac{\partial}{\partial w} J(w;\alpha) = w - \alpha x = 0 \tag{3}$$

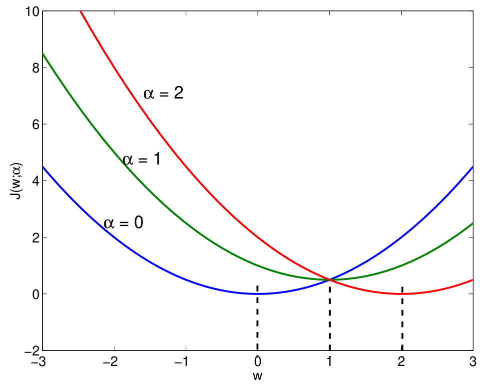

Figure 1: $J(w;\alpha)$ as a function of w for different values of α. The minimizing w values are indicated with dashed line segments. x was set to 1.

meaning that $w^*_\alpha = \alpha x$. We could now find the setting of α such that the contraint $w^*_\alpha x - 1 \geq 0$ is satisfied. There are multiple answers to this since larger values of α would better satisfy the constraint. Finding the smallest (non-negative) α for which the constraint is satisfied would in this case produce the right solution (one corresponding to the minimal change of the original problem).

We can proceed a bit more generally, however, the way we handled the quadratic optimization problem for SVMs. Let's insert our solution w^*_α back into the objective function:

$$J(w^*_\alpha;\alpha) = \frac{1}{2}(w^*_\alpha)^2 - \alpha(w^*_\alpha x - 1) = \frac{1}{2}(\alpha x)^2 - \alpha(\alpha x^2 - 1) = \alpha - \frac{1}{2}(\alpha x)^2 \tag{4}$$

The result, which we denote as $J(\alpha)$, is a function of the Lagrange multipler α only. Let's understand this function a bit better. In Figure 1, the values of the objective at the dashed lines correspond exactly to $J(w^*_\alpha;\alpha)$ or $J(\alpha)$, evaluated at $\alpha = 0, 1, 2$. Isn't it strange that the right solution ($\alpha = 1$) appears to yield the maximum of $J(\alpha)$? This is a very useful property. Let's verify this by finding the maximum of $J(\alpha)$ a bit more formally:

$$J(\alpha) = \alpha - \frac{1}{2}(\alpha x)^2 \tag{5}$$

$$\frac{\partial}{\partial \alpha} J(\alpha) = 1 - \alpha x^2 = 1 - w^*_a x = 0 \tag{6}$$

where we have used our previous result $w^*_\alpha = \alpha x$. So, the constraint is satisfied with equality at the maximum of $J(\alpha)$. More rigorously, since $\alpha \geq 0$ in our setting, the maximum is obtained either at $\alpha = 0$ or at the point where $1 - w^*_a x = 0$. We can express this more concisely by saying that their product vanishes, i.e., $\alpha(w^*_a x - 1) = 0$ at the optimum. This is generally true, i.e., either the Lagrange multiplier is not used and $\alpha = 0$ (the constraint is satisfied without any modification) or the Lagrange multiplier is positive and the constraint is satisfied with equality.

The remaining question for us here is why

$$\text{maximize } \alpha - \frac{1}{2}(\alpha x)^2 \text{ subject to } \alpha \geq 0 \tag{7}$$

is any better than the problem we started with. The short answer is that the constraints here are very simple non-negativity constraints that are easy to deal with in the optimization. In the SVM context, we have another reason to prefer this formulation.