Solving the 2-point sub problem

Note this writeup has an accompanying Mathematica notebook (screenshot) validating the algebra in this document is correct.

The original optimization problem is

$$\begin{aligned} \max_{\alpha} \quad & \sum_{j=1}^{m}\alpha_j - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle \\ \text{s.t.} \quad & 0 \le \alpha_i \le C, \\ & \sum_{i=1}^{m} y_i\alpha_i = 0 \end{aligned}$$

If we fix all but α_1, α_2 to constants and optimize for these two variables alone, we can rewrite the objective as

$$\begin{aligned} f(\alpha_1, \alpha_2) = \alpha_1 + \alpha_2 + \left[\sum_{j=3}^{m}\alpha_j\right] - \frac{1}{2}\Bigg[\\ & \alpha_1^2\langle x_1, x_1\rangle + \alpha_2^2\langle x_2, x_2\rangle + 2\alpha_1\alpha_2 y_1 y_2\langle x_1, x_2\rangle + \\ & 2\alpha_1 y_1 \sum_{i=3}^{m}\alpha_j y_j\langle x_1, x_j\rangle + \\ & 2\alpha_2 y_2 \sum_{i=3}^{m}\alpha_j y_j\langle x_2, x_j\rangle + \\ & \sum_{i=3}^{m}\sum_{j=3}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle \\ & \Bigg] \end{aligned}$$

Define

$$\begin{aligned} N_1 &= 2\sum_{i=3}^{m}\alpha_j y_j\langle x_1, x_j\rangle \\ N_2 &= 2\sum_{i=3}^{m}\alpha_j y_j\langle x_2, x_j\rangle \end{aligned}$$

Further note that, if we call $h(u) = \sum_{j=1}^{m}\alpha_j y_j\langle u, x_j\rangle$, i.e. the evaluation of the current separating hyperplane without taking the sign, then

$$
\begin{aligned}
N_1 &= h(x_1) - \alpha_1 y_1 \langle x_1, x_1 \rangle - \alpha_2 y_2 \langle x_1, x_2 \rangle \\
N_2 &= h(x_2) - \alpha_1 y_1 \langle x_2, x_1 \rangle - \alpha_2 y_2 \langle x_2, x_2 \rangle
\end{aligned}
$$

This will come in handy toward the end. Simplifying f appropriately and grouping all the constants into one term

$$
f(\alpha_1, \alpha_2) = \alpha_1 + \alpha_2 + O(1) - \frac{1}{2}\Bigg[\alpha_1 y_1 N_1 + \alpha_2 y_2 N_2 + \alpha_1^2 \langle x_1, x_1 \rangle + \alpha_2^2 \langle x_2, x_2 \rangle + 2\alpha_1 \alpha_2 y_1 y_2 \langle x_1, x_2 \rangle \Bigg]
$$

Now incorporating the linear constraint $\sum_{i=1}^m y_i \alpha_i = 0$, set P as follows, using the fact that $y_1 = 1/y_1$.

$$
\alpha_1 + \alpha_2 y_1 y_2 = P = -y_1 \sum_{j=3}^{m} \alpha_j y_j,
$$

and solve for α_1:

$$
\alpha_1 = (P - \alpha_2 y_1 y_2)
$$

Plug this into $f(\alpha_1, \alpha_2)$, define $g(\alpha_2)$ as the result, and group terms by like powers of α_2.

$$
\begin{aligned}
g(\alpha_2) &= f((P - \alpha_2 y_1 y_2), \alpha_2) \\
&= \left(-\frac{1}{2}\langle x_1, x_1 \rangle - \frac{1}{2}\langle x_2, x_2 \rangle + \langle x_1, x_2 \rangle \right) \alpha_2^2 + \\
&\quad (1 - N_2 y_2 - y_1 y_2 + P y_1 y_2 \langle x_1, x_2 \rangle - P y_1 y_2 \langle x_1, x_2 \rangle + N_1 y_2)\alpha_2 \\
&\quad + O(1)
\end{aligned}
$$

Now this is a single-variable polynomial we can optimize. Set

$$
\begin{aligned}
s &= 1 - N_2 y_2 - y_1 y_2 + P y_1 y_2 \langle x_1, x_2 \rangle - P y_1 y_2 \langle x_1, x_2 \rangle + N_1 y_2 \\
t &= -\frac{1}{2}\langle x_1, x_1 \rangle - \frac{1}{2}\langle x_2, x_2 \rangle + \langle x_1, x_2 \rangle
\end{aligned}
$$

This gives us $g(\alpha_2) = O(1) + s\alpha_2 + t\alpha_2^2$. We want to optimize this function using standard calculus. The first derivative is $g'(\alpha_2) = 2t\alpha_2 + s$ and vanishes at $\alpha_2 = -s/2t$. The second derivative is $2t = 2\langle x_1, x_2 \rangle - \|x_1\|^2 - \|x_2\|^2$. Recalling the elementary triangle inequality, which states that for vectors a, b

$$
0 \leq \|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\langle a, b \rangle
$$

It follows immediately that $2t < 0$ and so the optimum of g is a maximum, as desired.

Next we compute $-s/2t$, which is mostly regrouping s. First, substitute back into s the definition of $P = (\alpha_1 + \alpha_2 y_1 y_2)$ and use the fact that $y_i^2 = 1$.

$$\begin{aligned} s = 1 - N_2 y_2 - y_1 y_2 + N_1 y_2 + y_1 y_2 \alpha_1 \langle x_1, x_1 \rangle \\ - y_1 y_2 \alpha_1 \langle x_1, x_2 \rangle + \alpha_2 \langle x_1, x_1 \rangle - \alpha_2 \langle x_1, x_2 \rangle \end{aligned}$$

A small trick: replace the leading 1 with y_2^2 to allow us to factor later.

Next, substitute in the following and simplify:

$$\begin{aligned} N_1 &= h(x_1) - \alpha_1 y_1 \langle x_1, x_1 \rangle - \alpha_2 y_2 \langle x_1, x_2 \rangle \\ N_2 &= h(x_2) - \alpha_1 y_1 \langle x_2, x_1 \rangle - \alpha_2 y_2 \langle x_2, x_2 \rangle \end{aligned}$$

The result is

$$\begin{aligned} &[y_2^2 + h(x_1) y_2 - h(x_2) y_2 - y_1 y_2] \\ &+ [\alpha_2 \langle x_1, x_1 \rangle + \alpha_2 \langle x_2, x_2 \rangle - 2\alpha_2 \langle x_1, x_2 \rangle] \end{aligned}$$

Factor and rearrange the first bracketed term as $y_2(h(x_1) - y_1 + h(x_2) - y_2)$. The second is $\alpha_2(\langle x_1, x_1 \rangle + \langle x_2, x_2 \rangle - 2\langle x_1, x_2 \rangle)$. Set $\eta = \langle x_1, x_1 \rangle + \langle x_2, x_2 \rangle - 2\langle x_1, x_2 \rangle$

$$\begin{aligned} -s/2t &= \frac{y_2(h(x_1) - y_1 + h(x_2) - y_2) + \alpha_2 \eta}{\eta} \\ &= \alpha_2 + \frac{y_2(h(x_1) - y_1 + h(x_2) - y_2)}{\eta} \end{aligned}$$