A Dual of the Augmented-SVM

Starting with the Lagrangian in 7

$$\mathcal{L}(\mathbf{w},b,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{P}\alpha_i\left(y_i(\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b)-1\right) - \sum_{i\in\mathcal{I}_+}\beta_i\left(\mathbf{w}^T\mathrm{J}(\mathbf{x}_i)\hat{\mathbf{x}}_i+\xi_i\right) + C\sum_{i\in\mathcal{I}_+}\xi_i - \sum_{i\in\mathcal{I}_+}\mu_i\xi_i + \sum_{i=1}^{N}\gamma_i\mathbf{w}^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_i. \quad \text{(A.1)}$$

Setting its derivative w.r.t all the variables and Lagrange multipliers to zero, we get

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{P}\alpha_i y_i\boldsymbol{\phi}(\mathbf{x}_i) - \sum_{i\in\mathcal{I}_+}\beta_i\mathrm{J}(\mathbf{x}_i)\hat{\mathbf{x}}_i + \sum_{i=1}^{N}\gamma_i\mathrm{J}(\mathbf{x}^*)\mathbf{e}_i = 0.\\ \Rightarrow\mathbf{w} &= \sum_{i=1}^{P}\alpha_i y_i\boldsymbol{\phi}(\mathbf{x}_i) + \sum_{i\in\mathcal{I}_+}\beta_i\mathrm{J}(\mathbf{x}_i)\hat{\mathbf{x}}_i - \sum_{i=1}^{N}\gamma_i\mathrm{J}(\mathbf{x}^*)\mathbf{e}_i; \end{aligned}\quad \text{(A.2)}$$

$$\frac{\partial\mathcal{L}}{\partial b} = \sum_{i=1}^{P}\alpha_i y_i = 0; \quad \text{(A.3)}$$

$$\frac{\partial\mathcal{L}}{\partial\xi_i} = C - \beta_i - \mu_i = 0 \qquad \forall i\in\mathcal{I}_+. \quad \text{(A.4)}$$

Using A.4 with positivity constraints of the Lagrange multipliers β_i and μ_i we get

$$0\le\beta_i\le C \qquad \forall i\in\mathcal{I}_+. \quad \text{(A.5)}$$

Using A.2, A.3 and A.4 in A.1 we get the dual objective function to be maximized as

$$\hat{\mathcal{L}}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{i=1}^{P}\alpha_i - \frac{1}{2}\mathbf{w}^T\mathbf{w}. \quad \text{(A.6)}$$

Further expanding using A.2 we have

$$\begin{aligned}\mathbf{w}^T\mathbf{w} &= \left(\sum_{i=1}^{P}\alpha_i y_i\boldsymbol{\phi}(\mathbf{x}_i)^T + \sum_{i\in\mathcal{I}_+}\beta_i\hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T - \sum_{i=1}^{N}\gamma_i\mathbf{e}_i^T\mathrm{J}(\mathbf{x}^*)^T\right)\\ &\quad\left(\sum_{j=1}^{P}\alpha_j y_j\boldsymbol{\phi}(\mathbf{x}_j) + \sum_{j\in\mathcal{I}_+}\beta_j\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j - \sum_{j=1}^{N}\gamma_j\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\right)\\ &= \sum_{i=1}^{P}\left(\sum_{j=1}^{P}\alpha_i y_i\alpha_j y_j\boldsymbol{\phi}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_j) + \sum_{j\in\mathcal{I}_+}\alpha_i y_i\beta_j\boldsymbol{\phi}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j - \sum_{j=1}^{N}\alpha_i y_i\gamma_j\boldsymbol{\phi}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\right)\\ &\quad+\sum_{i\in\mathcal{I}_+}\left(\sum_{j=1}^{P}\beta_i\alpha_j y_j\hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_j) + \sum_{j\in\mathcal{I}_+}\beta_i\beta_j\hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j - \sum_{j=1}^{N}\beta_i\gamma_j\hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\right)\\ &\quad-\sum_{i=1}^{N}\left(\sum_{j=1}^{P}\gamma_i\alpha_j y_j\mathbf{e}_i^T\mathrm{J}(\mathbf{x}^*)^T\boldsymbol{\phi}(\mathbf{x}_j) + \sum_{j\in\mathcal{I}_+}\gamma_i\beta_j\mathbf{e}_i^T\mathrm{J}(\mathbf{x}^*)^T\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j - \sum_{j=1}^{N}\gamma_i\gamma_j\mathbf{e}_i^T\mathrm{J}(\mathbf{x}^*)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\right).\end{aligned}\quad \text{(A.7)}$$

Rewriting in matrix form,

$$\mathbf{w}^T\mathbf{w} = \begin{bmatrix}\boldsymbol{\alpha}^T & \boldsymbol{\beta}^T & \boldsymbol{\gamma}^T\end{bmatrix}\begin{bmatrix}\mathrm{K} & \mathrm{G} & -\mathrm{G}_* \\ \mathrm{G}^{\mathrm{T}} & \mathrm{H} & -\mathrm{H}_* \\ -\mathrm{G}_*^{\mathrm{T}} & -\mathrm{H}_*^{\mathrm{T}} & \mathrm{H}_{**}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{bmatrix} \quad \text{(A.8)}$$

where $\mathrm{K}\in\mathbb{R}^{P\times P}$, $\mathrm{G}\in\mathbb{R}^{|\mathcal{I}_+|\times P}$, $\mathrm{G}_*\in\mathbb{R}^{P\times N}$, $\mathrm{H}\in\mathbb{R}^{|\mathcal{I}_+|\times|\mathcal{I}_+|}$, $\mathrm{H}_*\in\mathbb{R}^{|\mathcal{I}_+|\times N}$, $\mathrm{H}_{**}\in\mathbb{R}^{N\times N}$ are given by

$$\left.\begin{array}{rlcrl}\mathrm{K}(i,j) &= y_i y_j\boldsymbol{\phi}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_j) & ; & \mathrm{H}(i,j) &= \hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j\\ \mathrm{G}(i,j) &= y_i\boldsymbol{\phi}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}_j)\hat{\mathbf{x}}_j & ; & \mathrm{H}_*(i,j) &= \hat{\mathbf{x}}_i^T\mathrm{J}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\\ \mathrm{G}_*(i,j) &= y_i\boldsymbol{\phi}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j & ; & \mathrm{H}_{**}(i,j) &= \mathbf{e}_i^T\mathrm{J}(\mathbf{x}^*)^T\mathrm{J}(\mathbf{x}^*)\mathbf{e}_j\end{array}\right\}. \quad \text{(A.9)}$$

Using Equations A.5, A.6 and A.8 the constrained optimization problem representing the dual can be now stated as

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} \frac{1}{2}\left[\boldsymbol{\alpha}^T\boldsymbol{\beta}^T\boldsymbol{\gamma}^T\right] \left[\begin{array}{ccc} \mathrm{K} & \mathrm{G} & -\mathrm{G}_* \\ \mathrm{G}^\mathrm{T} & \mathrm{H} & -\mathrm{H}_* \\ -\mathrm{G}_*^\mathrm{T} & -\mathrm{H}_*^\mathrm{T} & \mathrm{H}_{**} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{array}\right] - \boldsymbol{\alpha}^T\overline{\mathbf{1}}$$

subject to

$$\left.\begin{array}{cc} 0 \leq \alpha_i & \forall i = 1 \cdots P \\ 0 \leq \beta_i \leq C & \forall i \in \mathcal{I}_+ \\ \sum_{i=1}^{P} \alpha_i y_i = 0 & \end{array}\right\}. \tag{A.10}$$

Once the optimal solution for the above problem is obtained, the modulation function can be written as

$$\begin{aligned} h(\mathbf{x}) &= \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}) + b \\ &= \sum_{i=1}^{P} \alpha_i y_i \boldsymbol{\phi}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}) + \sum_{i\in\mathcal{I}_+} \beta_i \hat{\dot{\mathbf{x}}}_i^T \mathrm{J}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}) - \sum_{i=1}^{N} \gamma_i \mathbf{e}_i^T \mathrm{J}(\mathbf{x}^*)^T\boldsymbol{\phi}(\mathbf{x}) + b \quad \text{(Using A.2)} \\ &= \sum_{i=1}^{P} \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + \sum_{i\in\mathcal{I}_+} \beta_i \dot{\mathbf{x}}_i^T \frac{\partial k(\mathbf{x}, \mathbf{x}_i)}{\partial \mathbf{x}_i} - \sum_{i=1}^{N} \gamma_i \mathbf{e}_i^T \frac{\partial k(\mathbf{x}, \mathbf{x}^*)}{\partial \mathbf{x}^*} + b \quad \text{(Using B.1)} \end{aligned} \tag{A.11}$$

B Kernel Derivatives

For scalar variables x_i and any feature transformation $\boldsymbol{\phi} : \mathbb{R} \mapsto \mathbb{R}^F$ we define a valid Mercer kernel as $k(\mathbf{x}_i, \mathbf{x}_j) \equiv \boldsymbol{\phi}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_j)$. If $\prime$ denotes the derivative w.r.t the state variable, then the identities $\boldsymbol{\phi}'(x_i)^T\boldsymbol{\phi}(x_j) = \frac{\partial k(x_i,x_j)}{\partial x_i}$ and $\boldsymbol{\phi}'(x_i)^T\boldsymbol{\phi}'(x_j) = \frac{\partial^2 k(x_i,x_j)}{\partial x_i \partial x_j}$ follow directly from the definition of the kernel. We can rewrite these identities for vector variables $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^\mathbb{N}$ by taking the derivative w.r.t one of the components (say $n - th$) as $\left(\frac{\partial \boldsymbol{\phi}(\mathbf{x}_i)}{\partial \mathbf{x}(n)}\right)^T \boldsymbol{\phi}(\mathbf{x}_j) = \frac{\partial k(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{x}_i(n)}$. Expanding the first vector term we get

$$\Rightarrow \left[\frac{\partial \phi_1(\mathbf{x}_i)}{\partial \mathbf{x}(n)}, \frac{\partial \phi_2(\mathbf{x}_i)}{\partial \mathbf{x}(n)}, \cdots, \frac{\partial \phi_F(\mathbf{x}_i)}{\partial \mathbf{x}(n)}\right] \boldsymbol{\phi}(\mathbf{x}_j) = \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i(n)}.$$

Stacking the above equation in rows for $n = 1...N$, we get

$$\left[\begin{array}{cccc} \frac{\partial \phi_1(\mathbf{x}_i)}{\partial \mathbf{x}(1)} & \frac{\partial \phi_2(\mathbf{x}_i)}{\partial \mathbf{x}(1)} & \cdots & \frac{\partial \phi_F(\mathbf{x}_i)}{\partial \mathbf{x}(1)} \\ \frac{\partial \phi_1(\mathbf{x}_i)}{\partial \mathbf{x}(2)} & \frac{\partial \phi_2(\mathbf{x}_i)}{\partial \mathbf{x}(2)} & \cdots & \frac{\partial \phi_F(\mathbf{x}_i)}{\partial \mathbf{x}(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \phi_1(\mathbf{x}_i)}{\partial \mathbf{x}(N)} & \frac{\partial \phi_2(\mathbf{x}_i)}{\partial \mathbf{x}(N)} & \cdots & \frac{\partial \phi_F(\mathbf{x}_i)}{\partial \mathbf{x}(N)} \end{array}\right] \boldsymbol{\phi}(\mathbf{x}_j) = \left[\begin{array}{c} \frac{\partial k(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{x}_i(1)} \\ \frac{\partial k(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{x}_i(2)} \\ \vdots \\ \frac{\partial k(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{x}_i(N)} \end{array}\right]$$

$$\Rightarrow \mathrm{J}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_j) = \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i} \tag{B.1}$$

where J denotes the standard Jacobian matrix for a vector valued function. Similarly, by writing the derivatives w.r.t $(n, m) - th$ dimension and putting them as the corresponding element of a Hessian matrix we get

$$\mathrm{J}(\mathbf{x}_i)^T\mathrm{J}(\mathbf{x}_j) = \frac{\partial^2 k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i \partial \mathbf{x}_j}. \tag{B.2}$$

C Expansions for RBF Kernel

The above formulation is generic and can be applied to any kernel. Here we give the rbf kernel specific expressions for the block matrices in A.9.

$$
\begin{aligned}
\mathrm{K}(i,j) &= y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) = y_i y_j e^{-d\|\mathbf{x}_i - \mathbf{x}_j\|^2} \\
\mathrm{G}(i,j) &= y_i \left(\frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_j} \right)^T \hat{\mathbf{x}}_j = -2 d y_i e^{-d\|\mathbf{x}_i - \mathbf{x}_j\|^2} \left(\mathbf{x}_j - \mathbf{x}_i\right)^T \hat{\mathbf{x}}_j
\end{aligned}
$$

Replacing $\mathbf{x}_j$ by $\mathbf{x}^*$ in the above equation we get

$$
\begin{aligned}
\mathrm{G}_*(i,j) &= y_i \left(\frac{\partial k(\mathbf{x}_i, \mathbf{x}^*)}{\partial \mathbf{x}^*} \right)^T \mathbf{e}_j = -2 d y_i e^{-d\|\mathbf{x}_i - \mathbf{x}^*\|^2} \left(\mathbf{x}^* - \mathbf{x}_i\right)^T \mathbf{e}_j \\
\mathrm{H}(i,j) &= \hat{\mathbf{x}}_i^T \frac{\partial^2 k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \hat{\mathbf{x}}_j = \hat{\mathbf{x}}_i^T \left[\frac{\partial}{\partial \mathbf{x}_i} \left\{ -2 d e^{-d\|\mathbf{x}_i - \mathbf{x}_j\|^2} \left(\mathbf{x}_j - \mathbf{x}_i\right) \right\} \right] \hat{\mathbf{x}}_j \\
&= 2 d e^{d\|\mathbf{x}_i - \mathbf{x}_j\|^2} \left[\hat{\mathbf{x}}_i^T \hat{\mathbf{x}}_j - 2d \left\{ \hat{\mathbf{x}}_i^T (\mathbf{x}_i - \mathbf{x}_j) \right\} \left\{ (\mathbf{x}_i - \mathbf{x}_j)^T \hat{\mathbf{x}}_j \right\} \right].
\end{aligned}
$$

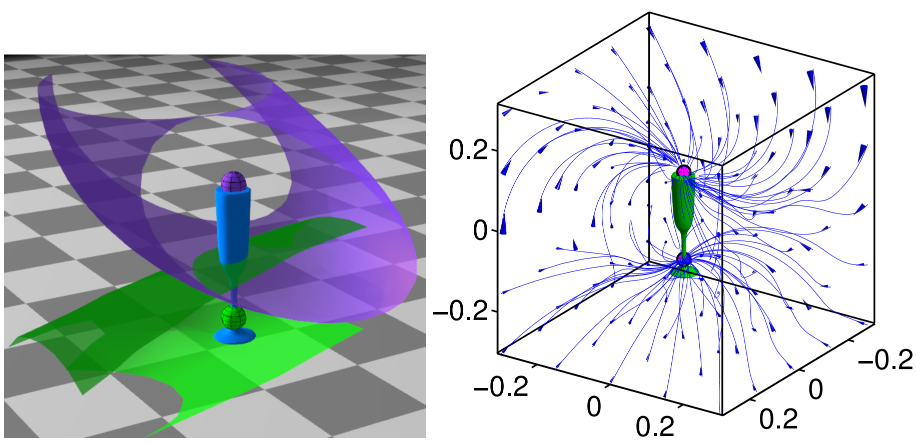
Again, replacing $\mathbf{x}_j$ by $\mathbf{x}^*$,

$$
\mathrm{H}_*(i,j) = \hat{\mathbf{x}}_i^T \frac{\partial^2 k(\mathbf{x}_i, \mathbf{x}^*)}{\partial \mathbf{x}_i \partial \mathbf{x}^*} \mathbf{e}_j = 2 d e^{d\|\mathbf{x}_i - \mathbf{x}^*\|^2} \left[\hat{\mathbf{x}}_i^T \mathbf{e}_j - 2d \left\{ \hat{\mathbf{x}}_i^T (\mathbf{x}_i - \mathbf{x}^*) \right\} \left\{ (\mathbf{x}_i - \mathbf{x}^*)^T \mathbf{e}_j \right\} \right].
$$

Replacing $\mathbf{x}_i$ also by $\mathbf{x}^*$,

$$
\mathrm{H}_{**}(i,j) = \mathbf{e}_i^T \frac{\partial^2 k(\mathbf{x}^*, \mathbf{x}^*)}{\partial \mathbf{x}^* \partial \mathbf{x}^*} \mathbf{e}_j = 2d \left(\mathbf{e}_i^T \mathbf{e}_j \right).
$$

D Champagne glass



(a) (b)

Figure D.1: (a) Two attractors placed on a champagne glass and their corresponding classification surfaces. (b) Complete flow of motion around the object. The supplementary video shows how the robot switches between these attractors to catch the glass in mid flight as it starts falling.