A Mathematical Details

Some mathematical expositions have been omitted in the main text to improve readability. Here we present this detailed material including the proofs of the statements.

A.1 Geometry in Pseudo-Euclidean Spaces

With this section, we want to contribute to a deeper understanding of the pE geometry by giving more details on the definitions given in Section 3.2.

With $\mathbb{R}^{(p,q)}$ we denote the pE space of signature (p, q), where $p, q \in \mathbb{N}_0$. This space can be seen as a product of a "real" and "imaginary" Euclidean vector space $\mathbb{R}^p \times i\mathbb{R}^q$, where $i = \sqrt{-1}$. Its elements are denoted with $\mathbf{z} = (\mathbf{z}_p^T, \mathbf{z}_q^T)^T$, the vector of real coordinates with respect to the basis $\{\mathbf{e}_k\}_{k=1}^p$ and $\{i\mathbf{e}_l\}_{l=1}^q$ of $\mathbb{R}^p$ and $i\mathbb{R}^q$. The *inner product* then is $\langle \mathbf{z}, \mathbf{z}' \rangle_{\mathrm{pE}} := \mathbf{z}_p^T \mathbf{z}_p' - \mathbf{z}_q^T \mathbf{z}_q' = \mathbf{z}^T \mathbf{M} \mathbf{z}'$ with $\mathbf{M} := \mathrm{diag}(\mathbf{1}_p, -\mathbf{1}_q)$. So this inner product is the difference of two standard Euclidean inner products. As the spaces are linear spaces, we can define *reduced convex hulls* of sets of points similar to [22, 23] as $\mathrm{conv}_\mu(\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}) := \{\sum_i \alpha_i \mathbf{z}_i | \sum_i \alpha_i = 1 \quad \text{and} \quad 0 \le \alpha_i \le \mu\}$. For $\mu = 1$ this is the usual convex hull. For smaller values μ the set reduces until, for $\mu = 1/n$, it consists of a single point, the mean of the $\mathbf{z}_i$. Fig. 5 illustrates such a sequence of (reduced) convex hulls.

Pseudo-Euclidean spaces in particular generalize Euclidean spaces by $q = 0$. We now focus on the case $q > 0$ and the corresponding differences in the resulting geometry. The inner product is a symmetric bilinear form as usual, but no longer positive definite. Nevertheless, similar to the Euclidean case we obtain the *squared norm* as $\|\mathbf{z}\|_{\mathrm{pE}}^2 := \langle \mathbf{z}, \mathbf{z} \rangle_{\mathrm{pE}}$. Note that this can be negative in contrast to the Euclidean case, thus it is not a norm in the strict sense. This notion immediately implies the *squared distance* of two points by $\|\mathbf{z}-\mathbf{z}'\|^2_{\mathrm{pE}} := \langle \mathbf{z}-\mathbf{z}', \mathbf{z}-\mathbf{z}' \rangle_{\mathrm{pE}} = \left\|\mathbf{z}_p - \mathbf{z}'_p\right\|^2 - \left\|\mathbf{z}_q - \mathbf{z}'_q\right\|^2$. *Orthogonality* is defined consequently by the inner product of two vectors being zero.

These definitions give rise to some interesting geometric phenomena. The first observation is that there are non-zero points which are orthogonal to themselves: $\langle \mathbf{z}, \mathbf{z} \rangle_{\mathrm{pE}} = 0$. These *isotropic* points form the *isotropic cone*, which separates the regions of points with positive and negative squared norm, cf. Fig. 6.

The squared distance between points is the difference of the corresponding squared distances in real and imaginary directions. This can be negative, so the real square root cannot necessarily be defined. Even if it could be defined, the resulting distance would not necessarily be a proper metric, as the triangle inequality does not have to hold. This is exactly the reason why arbitrary symmetric distances (without the requirement of metricity) can be represented in these spaces. Examples of squared distances are given in Fig. 7 a). The shaded region demonstrates a violation of the triangle inequality by $\sqrt{4} > \sqrt{0} + \sqrt{0}$.

The mapping $\mathbf{Mz}$ defines the reflection of a vector in $\mathbb{R}^{(p,q)}$ with respect to the real space $\mathbb{R}^p$. With this in mind, it is obvious that two vectors $\mathbf{z}, \mathbf{z}'$ are orthogonal, if the reflection $\mathbf{Mz}$ is orthogonal to $\mathbf{z}'$ in the Euclidean space $\mathbb{R}^{p+q}$.

Hyperplanes can be defined as usual: $H : \langle \mathbf{w}, \mathbf{z} \rangle_{\mathrm{pE}} + b = 0$. The normal vector $\mathbf{w}$ is orthogonal

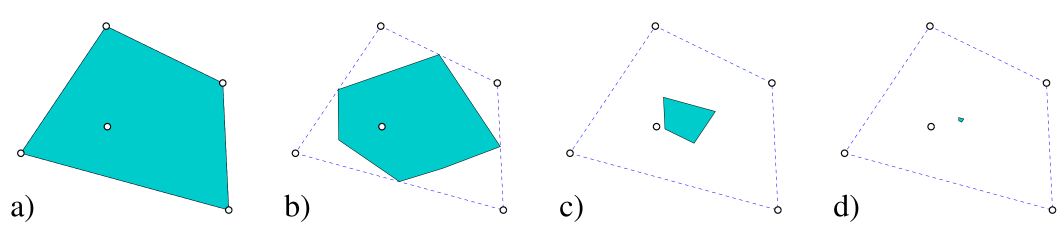

Figure 5: Illustration of (reduced) convex hulls for a set of 5 points. a) (Non-reduced) convex hull for $\mu = 1$, b) – d) reduced convex hulls for $\mu = 0.5, 0.3$ and 0.2.

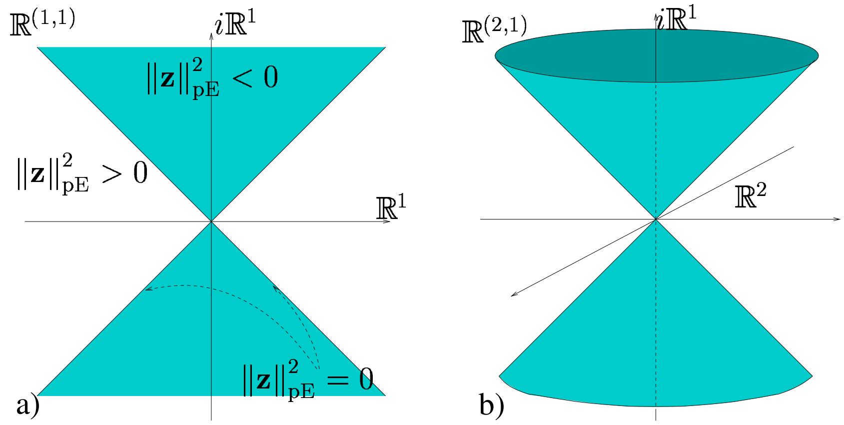

Figure 6: Illustration of pE spaces and their partition by isotropic cones into sets of positive, zero and negative squared norm. a) pE space $\mathbb{R}^{(1,1)}$, b) pE space $\mathbb{R}^{(2,1)}$.

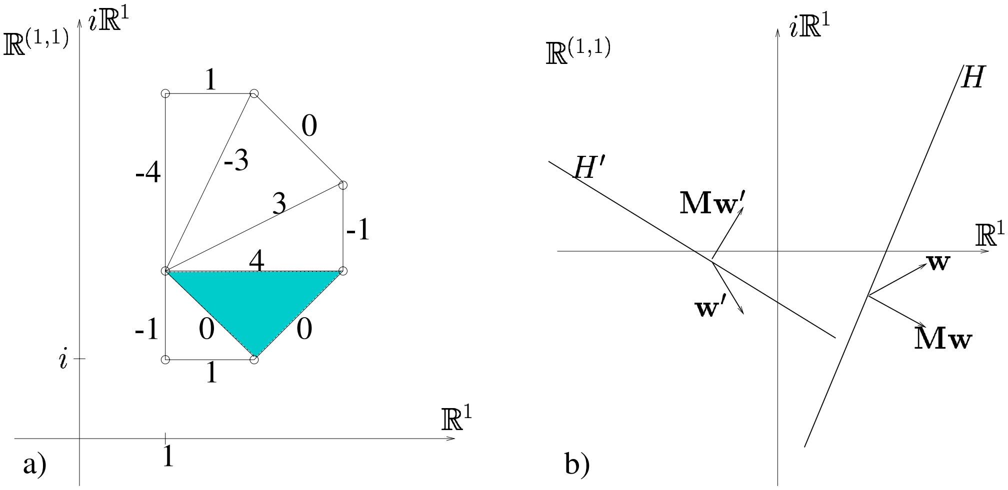

Figure 7: Geometry in pE spaces. a) Examples of squared distances between points and a violation of the triangle inequality (shaded), b) hyperplanes H, H' and corresponding pE normals $\mathbf{w}$, $\mathbf{w}'$, the flipped versions $\mathbf{M}\mathbf{w}$, $\mathbf{M}\mathbf{w}'$ of which are Euclidean normals.

to the plane H in the pE sense. Linear classification can easily be performed by taking the sign of any such linear function. Fig. 7 b) illustrates these aspects.

A.2 Isometric Embedding

We continue with the constructive proof of Prop. 1 and conclude with some useful relations between the geometric functions.

Proof of Proposition 1 (Isometric Embedding). We use the construction from [19]. The symmetric and zero-diagonal function d^2 allows to define the matrix of squared distances by $\mathbf{D}^{(2)} := (d^2(x_i, x_j))_{i,j=1}^n$. With the centering matrix $\mathbf{J} := \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ we can construct the *centered* kernel matrix $\mathbf{K} := -\frac{1}{2}\mathbf{J}\mathbf{D}^{(2)}\mathbf{J}$. This matrix is symmetric and singular, as $\mathbf{J}$ has eigenvalue 0. The eigendecomposition can be performed as $\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ with $\mathbf{U}$ being orthogonal and $\mathbf{\Lambda}$ being diagonal starting with the p positive eigenvalues, followed by the q negative ones and $n - p - q \geq 1$ eigenvalues equal 0. Therefore $p + q < n$ and

$$\mathbf{K} = \mathbf{U}|\mathbf{\Lambda}|^{1/2}\mathrm{diag}(\mathbf{1}_p, -\mathbf{1}_q, \mathbf{0}_{n-p-q})|\mathbf{\Lambda}|^{1/2}\mathbf{U}^T.$$

This can be interpreted as an inner product matrix in the pE space $\mathbb{R}^{(p,q)}$: Using the corresponding matrix $\mathbf{M} := \mathrm{diag}(\mathbf{1}_p, -\mathbf{1}_q)$ and defining the mapping $\Phi : \{x_i\}_{i=1}^n \to \mathbb{R}^{(p,q)}$ by $\Phi(x_i)$ being the vector consisting of the first $p+q$ components of the i-th column of $|\mathbf{\Lambda}|^{1/2}\mathbf{U}^T$ we obtain the matrix element $K_{ij} = \Phi(x_i)^T\mathbf{M}\Phi(x_j) = \langle\Phi(x_i), \Phi(x_j)\rangle_{\mathrm{pE}}$.

As $\|\Phi(x_i) - \Phi(x_j)\|_{\mathrm{pE}}^2 = K_{ii} - 2K_{ij} + K_{jj}$ it suffices for the isometry to show that the right hand is equal to $d^2(x_i, x_j)$. This follows by simple calculations from the fact that the elements of $\mathbf{K}$ are defined as

$$K_{kl} = -\frac{1}{2}\left(d^2(x_k, x_l) - \frac{1}{n}\sum_i d^2(x_k, x_i) - \frac{1}{n}\sum_i d^2(x_i, x_l) + \frac{1}{n^2}\sum_{i,j} d^2(x_i, x_j)\right).$$

□

For the proofs of the theoretical statements and various reformulations throughout the text the following relations are applied. They state that the kernel k, the induced squared distance function $-\frac{1}{2}d^2$ and the inner product $\langle\cdot,\cdot\rangle_{\mathrm{pE}}$ of any embedding pE space essentially are identical. They only differ by some suitable function in one variable.

Lemma 6 (Relation of Geometric Functions). *For any embedding of data according to Prop. 1, some suitable functions h and h' exist such that for all i, j holds*

$$k(x_i, x_j) = -\frac{1}{2}d^2(x_i, x_j) + h'(x_i) + h'(x_j) = \Phi(x_i)^T\mathbf{M}\Phi(x_j) + h(x_i) + h(x_j). \tag{10}$$

This particularly implies for all y_i and α_i with $\sum_i y_i\alpha_i = 0$

$$\sum_{i,j}\alpha_i\alpha_j y_i y_j k(x_i, x_j) = -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j d^2(x_i, x_j) = \sum_{i,j}\alpha_i\alpha_j y_i y_j \Phi(x_i)^T\mathbf{M}\Phi(x_j). \tag{11}$$

Proof of Lemma 6. The first equality in (10) follows from (1) with $h'(x) := \frac{1}{2}k(x, x)$. The second follows from (2) by decomposing the pE norm in inner products and setting $h(x) := h'(x) + \frac{1}{2}\Phi(x_i)^T\mathbf{M}\Phi(x_j)$. Equation (11) then is a simple implication, since the terms containing h or h' cancel out due to $\sum_i y_i\alpha_i = 0$. □

A.3 Derivation of CH Classification

We want to give details on the derivation of (CH-DU) in Section 4 when starting from (4): In analogy to the Euclidean case [29], $\mathbf{z}^-$ and $\mathbf{z}^+$ are convex combinations, i.e.

$$\mathbf{z}^\pm = \sum_{i:y_i=\pm1} \bar{\alpha}_i \Phi(x_i) \quad \text{with} \quad \sum_{i:y_i=+1} \bar{\alpha}_i = 1, \quad \sum_{i:y_i=-1} \bar{\alpha}_i = 1 \quad \text{and} \quad 0 \leq \bar{\alpha}_i, \tag{12}$$

which implies $\sum_i \bar{\alpha}_i = 2$ and $\sum_i y_i \bar{\alpha}_i = 0$. So the constraints of (CH-DU) are obtained. Using these representations, bilinearity of the inner product and (11) from Lemma 6 yields for the objective function of (4):

$$\begin{aligned} \left\|\mathbf{z}^+ - \mathbf{z}^-\right\|_{\mathrm{pE}}^2 &= \left\|\sum_i \bar{\alpha}_i y_i \Phi(x_i)\right\|_{\mathrm{pE}}^2 = \sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i y_j \left\langle \Phi(x_i), \Phi(x_j) \right\rangle_{\mathrm{pE}} \\ &= -\frac{1}{2} \sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i y_j d^2(x_i, x_j). \end{aligned}$$

This is exactly the (negative) of the objective function of (CH-DU), which concludes the derivation of (CH-DU).

We continue with the derivation of the CH classification rule (7) starting from (6): Using definitions yields

$$\begin{aligned} g(\mathbf{z}) &= \left\|\mathbf{z} - \sum_{i:y_i=-1} \bar{\alpha}_i \Phi(x_i)\right\|_{\mathrm{pE}}^2 - \left\|\mathbf{z} - \sum_{i:y_i=+1} \bar{\alpha}_i \Phi(x_i)\right\|_{\mathrm{pE}}^2 \\ &= -2 \sum_{i:y_i=-1} \bar{\alpha}_i \left\langle \mathbf{z}, \Phi(x_i) \right\rangle_{\mathrm{pE}} + 2 \sum_{i:y_i=+1} \bar{\alpha}_i \left\langle \mathbf{z}, \Phi(x_i) \right\rangle_{\mathrm{pE}} \\ &\quad + \sum_{i,j:y_i=y_j=-1} \bar{\alpha}_i \bar{\alpha}_j \left\langle \Phi(x_i), \Phi(x_j) \right\rangle_{\mathrm{pE}} - \sum_{i,j:y_i=y_j=+1} \bar{\alpha}_i \bar{\alpha}_j \left\langle \Phi(x_i), \Phi(x_j) \right\rangle_{\mathrm{pE}} \end{aligned}$$

$$
\begin{aligned}
&= +2\sum_i \bar{\alpha}_i y_i \langle \mathbf{z}, \Phi(x_i)\rangle_{\mathrm{pE}} - \sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \langle \Phi(x_i), \Phi(x_j)\rangle_{\mathrm{pE}} \\
&=: T_1 + T_2.
\end{aligned}
$$

Rewriting both terms using $\langle \mathbf{z}, \mathbf{z}' \rangle_{\mathrm{pE}} = -\frac{1}{2}\left(\|\mathbf{z}-\mathbf{z}'\|_{\mathrm{pE}}^2 - \|\mathbf{z}\|_{\mathrm{pE}}^2 - \|\mathbf{z}'\|_{\mathrm{pE}}^2\right)$ yields

$$
\begin{aligned}
T_1 &= -\sum_i \bar{\alpha}_i y_i \|\mathbf{z} - \Phi(x_i)\|_{\mathrm{pE}}^2 + \sum_i \bar{\alpha}_i y_i \|\Phi(x_i)\|_{\mathrm{pE}}^2 \\
T_2 &= \frac{1}{2}\sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \|\Phi(x_i) - \Phi(x_j)\|_{\mathrm{pE}}^2 - \frac{1}{2}\sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \|\Phi(x_i)\|_{\mathrm{pE}}^2 - \frac{1}{2}\sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \|\Phi(x_j)\|_{\mathrm{pE}}^2 \\
&= \frac{1}{2}\sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \|\Phi(x_i) - \Phi(x_j)\|_{\mathrm{pE}}^2 - \sum_i \bar{\alpha}_i y_i \|\Phi(x_i)\|_{\mathrm{pE}}^2 + 0.
\end{aligned}
$$

The last reformulation follows from the constraints of (CH-DU). Obviously, the sum of those both terms results in the desired equation

$$
g(\mathbf{z}) = -\sum_i \bar{\alpha}_i y_i \|\mathbf{z} - \Phi(x_i)\|_{\mathrm{pE}}^2 + \frac{1}{2}\sum_{i,j} \bar{\alpha}_i \bar{\alpha}_j y_i \|\Phi(x_i) - \Phi(x_j)\|_{\mathrm{pE}}^2. \tag{13}
$$

In particular this implies that if a point $\mathbf{z}$ has a pre-image $x \in \mathcal{X}$, i.e. $\mathbf{z} = \Phi(x)$, we can perform the classification in the original space without explicit mapping by taking the sign of (7).

A.4 CH Primal Optimization Problem

We continue with the primal optimization problem of the CH classifier. Note that the choice of b differs from (7), but we want to keep the standard notation. This result is an extension of the Euclidean case in [23].

Proposition 7 (CH Primal in $\mathbb{R}^{(p,q)}$). *Let $\bar{\boldsymbol{\alpha}}$ be a stationary point of (CH-DU), such that there* *exist two non-bounded coefficients of different classes, i.e.* $0 < \bar{\alpha}_k, \bar{\alpha}_l < \mu$ *with* $y_k = +1$ *and* $y_l = -1$*, which induce a positive* ρ *as defined below. Let* $\Phi : \{x_i\}_{i=1}^n \to \mathbb{R}^{(p,q)}$ *be an isometric embedding according to Proposition 1.*

Then we obtain a stationary point $\mathbf{w} \in \mathbb{R}^{(p,q)}, b \in \mathbb{R}, \rho \in \mathbb{R}_+$ *and* $\boldsymbol{\xi} \in \mathbb{R}_+^n$ *of the convex hull primal optimization problem*

$$\begin{aligned} \min_{\mathbf{w},b,\rho,\boldsymbol{\xi}} \quad & \tfrac{1}{2}\mathbf{w}^T\mathbf{M}\mathbf{w} - 2\rho + \mu \textstyle\sum_i \xi_i \\ \textit{s.t.} \quad & y_i(\mathbf{w}^T\mathbf{M}\Phi(x_i) + b) \geq \rho - \xi_i, \quad \xi_i \geq 0 \quad \textit{and} \quad \rho \geq 0 \end{aligned}$$

by setting $\mathbf{w} = \sum_i \bar{\alpha}_i y_i \Phi(x_i)$, $b := -\frac{1}{2}\left(\mathbf{w}^T\mathbf{M}\Phi(x_k) + \mathbf{w}^T\mathbf{M}\Phi(x_l)\right)$,

$$\rho := \frac{1}{2}\mathbf{w}^T\mathbf{M}(\Phi(x_k) - \Phi(x_l))$$

and

$$\xi_i := \begin{cases} \rho - y_i(\mathbf{w}^T\mathbf{M}\Phi(x_i) + b) & \textit{if} \quad \bar{\alpha}_i = \max_j \bar{\alpha}_j \\ 0 & \textit{otherwise.} \end{cases}$$

It can be proven by the stationarity of $\bar{\boldsymbol{\alpha}}$ that x_k and x_l can be chosen arbitrarily from both sets of unbounded samples which implies well-definedness of b and ρ. An exact correspondence of local optima can also be achieved under some more restrictive conditions, namely requiring that for all feasible directions $(\Delta\mathbf{w}^T, \Delta b, \Delta\rho, \Delta\boldsymbol{\xi}^T)^T$ in which the derivative of the primal optimization function vanishes, the inequality $\Delta\mathbf{w}^T\mathbf{M}\Delta\mathbf{w} \geq 0$ holds.

Proof of Proposition 7 (CH Primal in $\mathbb{R}^{(p,q)}$*).* We start with an equivalence that will be used, where

$\mathbf{B}$ denotes the matrix with entries $B_{ij} := -\frac{1}{2}d^2(x_i, x_j)y_iy_j$: For all i, j holds

$$\mathbf{w}^T\mathbf{M}(\Phi(x_i) - \Phi(x_j)) = \bar{\boldsymbol{\alpha}}^T\mathbf{B}(y_i\mathbf{e}_i - y_j\mathbf{e}_j). \tag{14}$$

This can be seen by expressing both sides in terms of squared distances. For the right hand we obtain by definition of $\mathbf{B}$

$$\bar{\boldsymbol{\alpha}}^T\mathbf{B}(y_i\mathbf{e}_i - y_j\mathbf{e}_j) = -\frac{1}{2}\sum_m \bar{\alpha}_m y_m(d^2(x_i, x_m) - d^2(x_j, x_m)).$$

For the left hand we can apply the second equality of (10) and obtain with $h''(x) := h'(x) - h(x)$

$$\mathbf{w}^T\mathbf{M}(\Phi(x_i) - \Phi(x_j)) = \sum_m \bar{\alpha}_m y_m \left(-\frac{1}{2}d^2(x_i, x_m) + \frac{1}{2}d^2(x_j, x_m) + h''(x_i) - h''(x_j)\right). \tag{15}$$

The last terms cancel out due to $\bar{\boldsymbol{\alpha}}^T\mathbf{y} = 0$, and we get the claimed identity (14).

We start with the proof of $(\mathbf{w}^T, b, \rho, \boldsymbol{\xi}^T)^T$ being a feasible point. Obviously $\mathbf{w}$ and b are not explicitly constrained and $\rho > 0$ by assumption.

We now argue why $\xi_i \geq 0$. If $\bar{\alpha}_i < \max_j \bar{\alpha}_j$ then $\xi_i = 0$ by definition. If $\alpha_i = \max_j \bar{\alpha}_j$ we have to show that ξ_i is non-negative. We assume $y_i = 1$, the other case follows similarly. Recall that $y_k = +1$ by assumption. By using the definitions of ρ and b and applying (14) we get

$$\begin{aligned}
\xi_i &= \rho - \frac{1}{2}\mathbf{w}^T\mathbf{M}(\Phi(x_i) - \Phi(x_k)) - \frac{1}{2}\mathbf{w}^T\mathbf{M}(\Phi(x_i) - \Phi(x_l)) \\
&= \rho - \mathbf{w}^T\mathbf{M}(\Phi(x_i) - \Phi(x_k)) - \frac{1}{2}\mathbf{w}^T\mathbf{M}(\Phi(x_k) - \Phi(x_l)) \\
&= \mathbf{w}^T\mathbf{M}(\Phi(x_k) - \Phi(x_i)) = \bar{\boldsymbol{\alpha}}^T\mathbf{B}(y_k\mathbf{e}_k - y_i\mathbf{e}_i) = \left\langle \bar{\boldsymbol{\alpha}}^T\mathbf{B}, \mathbf{e}_k - \mathbf{e}_i \right\rangle \geq 0.
\end{aligned}$$

The last step follows from $\Delta\bar{\boldsymbol{\alpha}} := \mathbf{e}_k - \mathbf{e}_i$ being a feasible direction of (CH-DU) and the derivative of (CH-DU) $\left\langle -\bar{\boldsymbol{\alpha}}^T\mathbf{B}, \Delta\bar{\boldsymbol{\alpha}}\right\rangle \leq 0$ due to the stationarity of $\bar{\boldsymbol{\alpha}}$. Feasibility of $\Delta\bar{\boldsymbol{\alpha}}$ can be seen since the feasible directions of (CH-DU) are characterized by

$$\mathbf{y}^T\Delta\bar{\boldsymbol{\alpha}} = 0, \quad \mathbf{1}^T\Delta\bar{\boldsymbol{\alpha}} = 0, \quad \Delta\bar{\alpha}_i \geq 0 \ \text{ if } \ \bar{\alpha}_i = 0 \quad \text{and} \quad \Delta\bar{\alpha}_i \geq 0 \ \text{ if } \ \bar{\alpha}_i = \mu. \tag{16}$$

We now show that the last constraint is satisfied, i.e. $y_i(\mathbf{w}^T\mathbf{M}\Phi(x_i) + b) \geq \rho - \xi_i$ holds for all i. In the case of $\bar{\alpha}_i = \max_j \bar{\alpha}_j$ this is valid with equality due to the definition of the ξ_i. In the case of $\bar{\alpha} < \max_j \bar{\alpha}_j$ we know $\xi_i = 0$ and it suffices to show that

$$y_i(\mathbf{w}^T\mathbf{M}\Phi(x_i) + b) - \rho \geq 0. \tag{17}$$

The statement follows similarly as above by using the definitions of ρ and b, (14) and the stationarity of $\bar{\boldsymbol{\alpha}}$. More precisely, (17) holds with equality if $0 < \bar{\alpha}_i < \max_j \bar{\alpha}_j$ and with "$\geq$" in the case of $\bar{\alpha}_i = 0$. Thus $(\mathbf{w}^T, b, \rho, \boldsymbol{\xi}^T)^T$ is a feasible point.

We continue with arguing that it is a stationary point. It is sufficient to show that for every feasible direction $\mathbf{v} := (\Delta\mathbf{w}^T, \Delta b, \Delta\rho, \Delta\boldsymbol{\xi}^T)^T$ of the primal optimization problem $J(\mathbf{w}, \Delta b, \Delta\rho, \boldsymbol{\xi})$ satisfies $\langle \nabla J, \mathbf{v}\rangle \geq 0$. By computing and inserting the partial derivatives, this is equivalent to showing

$$\mathbf{w}^T\mathbf{M}\Delta\mathbf{w} + \mu\mathbf{1}^T\Delta\boldsymbol{\xi} - 2\Delta\rho \geq 0. \tag{18}$$

If $\alpha_i > 0$ then the slack constraint is satisfied with equality, so a feasible direction $\mathbf{v}$ satisfies

$$y_i(\Delta\mathbf{w}^T\mathbf{M}\Phi(x_i)) \geq \Delta\rho - \Delta\xi_i - y_i\Delta b.$$

Multiplying with $\bar{\alpha}_i$, summing over all i, using $\sum_i \bar{\alpha}_i = 2$ and the definition of $\mathbf{w}$ we obtain

$$\mathbf{w}^T\mathbf{M}\Delta\mathbf{w} + \bar{\boldsymbol{\alpha}}^T\Delta\boldsymbol{\xi} - 2\Delta\rho \geq 0.$$

So (18) is particularly satisfied as $\mu \geq \bar{\alpha}_i$ for all i. We conclude that $(\mathbf{w}^T, b, \rho, \boldsymbol{\xi}^T)^T$ is a stationary point of the primal optimization problem. □

A.5 Equivalence of CH and SVM

Here we present the detailed proof of the main result, the relation between CH and SVM classification.

Proof of Proposition 3 (Equivalence of CH and SVM). (SVM-DU) is equivalent to

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \mathbf{1}_n^T\boldsymbol{\alpha} - \tfrac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} \\ \text{s.t.} \quad & \mathbf{0}_n \leq \boldsymbol{\alpha} \leq C\mathbf{1}_n \quad \text{and} \quad \mathbf{y}^T\boldsymbol{\alpha} = 0. \end{aligned} \tag{19}$$

Similarly, with Lemma 6, (CH-DU) is equivalent to

$$\begin{aligned} \max_{\bar{\boldsymbol{\alpha}}} \quad & -\tfrac{1}{2}\bar{\boldsymbol{\alpha}}^T\mathbf{Q}\bar{\boldsymbol{\alpha}} \\ \text{s.t.} \quad & \mathbf{0}_n \leq \bar{\boldsymbol{\alpha}} \leq \mu\mathbf{1}_n, \quad \mathbf{y}^T\bar{\boldsymbol{\alpha}} = 0 \quad \text{and} \quad \mathbf{1}_n^T\bar{\boldsymbol{\alpha}} = 2. \end{aligned} \tag{20}$$

So it is sufficient to show the correspondences of local solutions of (19) and (20).

A feasible direction $\mathbf{v}$ of (SVM-DU) in the point $\boldsymbol{\alpha}$ is characterized by

$$\mathbf{y}^T\mathbf{v} = 0, \quad v_i \geq 0 \quad \text{if} \quad \alpha_i = 0 \quad \text{and} \quad v_i \geq 0 \quad \text{if} \quad \alpha_i = C. \tag{21}$$

Similarly but with an additional constraint, we already characterized the feasible directions of (CH-DU) in (16).

i) A non-zero local optimum $\boldsymbol{\alpha}$ of (19) clearly induces a feasible point $\bar{\boldsymbol{\alpha}}$ of (20), as the scaling exactly results in $\mathbf{1}^T\boldsymbol{\alpha} = 2$ and corresponding bounding constraints. It remains to check that $\bar{\boldsymbol{\alpha}}$ is a stationary point and even a local optimum. Let $\mathbf{v}$ be a feasible direction of (CH-DU). As $\mathbf{v}$ simultaneously is a feasible direction of (SVM-DU) and $\boldsymbol{\alpha}$ a stationary point we know $\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, \mathbf{v}\rangle \leq 0$. Since $\mathbf{1}_n^T\mathbf{v} = 0$ we get non-positive derivatives of (20): $\langle -\mathbf{Q}\bar{\boldsymbol{\alpha}}, \mathbf{v}\rangle \leq 0$. So $\bar{\boldsymbol{\alpha}}$ is a stationary point of (CH-DU). It even is a local optimum: Let $\mathbf{v}$ be a feasible direction of (CH-DU) with $\langle -\mathbf{Q}\bar{\boldsymbol{\alpha}}, \mathbf{v}\rangle = 0$. Then we get $\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, \mathbf{v}\rangle = 0$. As $\mathbf{v}$ is a feasible direction of (SVM-DU), the second derivative is non-positive $-\mathbf{v}^T\mathbf{Q}\mathbf{v} \leq 0$. The second derivatives of (SVM-DU) and (CH-DU) coincide, so we have shown that the curvature of (CH-DU) in the direction $\mathbf{v}$ is non-positive and $\bar{\boldsymbol{\alpha}}$ is a local optimum.

ii) Clearly a local solution $\bar{\boldsymbol{\alpha}}$ of (20) induces a feasible point $\boldsymbol{\alpha}$ of (19) for any choice of $\rho > 0$. It remains to show that it is a stationary point and a local optimum.

For being a stationary point we show that for every feasible direction $\mathbf{v}$ satisfying (21) holds

$$\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, \mathbf{v}\rangle \leq 0. \tag{22}$$

We have two unbounded coefficients α_k, α_l with $y_k = +1$ and $y_l = -1$. Then we decompose $\mathbf{v}$:

$$\begin{aligned}
\mathbf{v} &= \sum_i v_i \mathbf{e}_i \\
&= \sum_{i:y_i=1} y_i v_i (y_i \mathbf{e}_i - y_k \mathbf{e}_k) + \sum_{i:y_i=-1} y_i v_i (y_i \mathbf{e}_i - y_l \mathbf{e}_l) + \sum_{i:y_i=1} v_i (y_k \mathbf{e}_k - y_l \mathbf{e}_l) \\
&=: T_1 + T_2 + T_3.
\end{aligned}$$

So it is sufficient to show (22) for contributions $\mathbf{v}'$ contained in T_1, T_2 or T_3. We start with T_1. For $\mathbf{v}' = y_i v_i (y_i \mathbf{e}_i - y_k \mathbf{e}_k)$ we obtain

$$\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, \mathbf{v}' \rangle = y_i v_i \langle \mathbf{1}_n, \mathbf{e}_i - \mathbf{e}_k \rangle - y_i v_i \frac{1}{\rho} \langle \mathbf{Q}\bar{\boldsymbol{\alpha}}, y_i \mathbf{e}_i - y_k \mathbf{e}_k \rangle . \tag{23}$$

The first term vanishes, and it remains to argue why the last term is non-positive. In the case of $0 < \alpha_i < C$, we know that $\bar{\alpha}_i$ due to scaling also is not bounded. So $\pm(y_i \mathbf{e}_i - y_k \mathbf{e}_k)$ are feasible directions of (CH-DU), and the corresponding derivatives $\pm \langle \mathbf{Q}\bar{\boldsymbol{\alpha}}, y_i \mathbf{e}_i - y_k \mathbf{e}_k \rangle \leq 0$, therefore the last term in (23) is 0. In the case of $\alpha_i = 0$, the corresponding $v_i \geq 0$ as $\mathbf{v}$ is a feasible direction of (SVM-DU). So $\mathbf{e}_i - y_i y_k \mathbf{e}_k$ is a feasible direction of (CH-DU), which implies that the last inner product in (23) has the same sign as y_i. We conclude similarly in the case of $\alpha_i = C$ that $v_i \leq 0$ and $-(\mathbf{e}_i - y_i y_k \mathbf{e}_k)$ is feasible for (CH-DU) which again yields the desired non-positivity of the last term of (23). For directions $\mathbf{v}'$ in T_2 the argumentation is analogous, so we continue with T_3. The vector $y_k \mathbf{e}_k - y_l \mathbf{e}_l$ is not a feasible direction of (CH-DU), so the argumentation is different. Similarly as before we have to show the non-positivity of

$$\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, v_i(\mathbf{e}_k + \mathbf{e}_l) \rangle = v_i \left(\langle \mathbf{1}_n, \mathbf{e}_k + \mathbf{e}_l \rangle - \frac{1}{\rho} \langle \mathbf{Q}\bar{\boldsymbol{\alpha}}, \mathbf{e}_k + \mathbf{e}_l \rangle \right) .$$

This is particularly satisfied, if the term in brackets vanishes, i.e. if $\rho := \frac{1}{2}\langle \mathbf{Q}\bar{\boldsymbol{\alpha}},\ \mathbf{e}_k + \mathbf{e}_l\rangle$ and if $\rho > 0$. So we conclude that $\boldsymbol{\alpha}$ is a stationary point of (SVM-DU).

The positive definiteness of $\mathbf{Q}$ on the space of all feasible directions $\mathbf{v}$ of (SVM-DU) with $\langle \mathbf{1}_n - \mathbf{Q}\boldsymbol{\alpha}, \mathbf{v}\rangle = 0$ further implies that $\boldsymbol{\alpha}$ is a local optimum.

iii) By expressing d^2 in terms of k as given in (10), equation (7) turns exactly into a (positively) scaled but possibly shifted version of (8). So the resulting hyperplanes are parallel.

In the case of no bounded coefficients $\bar{\boldsymbol{\alpha}}$ (or $\boldsymbol{\alpha}$) one can show that for two (trivially existing) unbounded $\bar{\alpha}_k$ and $\bar{\alpha}_l$ of different classes $y_k = +1$ and $y_l = -1$ the equation $f(x_k) = -f(x_l)$ holds, if f is chosen as the SVM (8) or CH classification (7) function. This implies the identity of the classification boundaries and regions. We demonstrate it for the CH classification function, for the SVM function it is analogous.

In the case of CH classification we already know that $g(\mathbf{z})$ in (6) has identical values for $\mathbf{z}^+$ and $\mathbf{z}^-$. So it remains to show that $f(x_k) = g(\mathbf{z}^+)$ and $f(x_l) = g(\mathbf{z}^-)$. We only show the first equality, the second follows analogously.

Using (13) and (7) we get after expressing d^2 by inner products by (10)

$$g(\mathbf{z}^+) - f(x_k) = \sum_i \bar{\alpha}_i \Phi(x_i)^T \mathbf{M} \left(\sum_{j:y_j=+1} \bar{\alpha}_j \Phi(x_j) - \Phi(x_k) \right).$$

Using the definition of $\mathbf{w}$ from Proposition 7 and $\sum_{j:y_j=+1} \bar{\alpha}_j = 1$ this is

$$= \mathbf{w}^T \mathbf{M} \left(\sum_{j:y_j=+1} \bar{\alpha}_j (\Phi(x_j) - \Phi(x_k)) \right).$$

In order to show that this is zero it is sufficient to show that all single terms in the sum vanish. By

(14) we get

$$\mathbf{w}^T \mathbf{M}(\Phi(x_j) - \Phi(x_k)) = \bar{\boldsymbol{\alpha}}^T \mathbf{B}(y_j \mathbf{e}_j - y_k \mathbf{e}_k).$$

This is exactly the derivative of (CH-DU) in the direction $\Delta\bar{\boldsymbol{\alpha}} := -\mathbf{e}_j + \mathbf{e}_k$. Since $\pm\Delta\bar{\boldsymbol{\alpha}}$ are feasible directions this must be 0 by stationarity of $\bar{\boldsymbol{\alpha}}$. □

A.6 Uniqueness of Stationary Points

We conclude with the last proof, the derivation of the uniqueness statement.

Proof of Lemma 5 (Uniqueness of Stationary Points). Let $(\bar{\mathbf{z}}^-, \bar{\mathbf{z}}^+)$ be another stationary point of (4). We denote the optimization function as J. Then $(\Delta\mathbf{z}^-, \Delta\mathbf{z}^+) := (\mathbf{z}^- - \bar{\mathbf{z}}^-, \mathbf{z}^+ - \bar{\mathbf{z}}^+)$ is a feasible direction in $(\bar{\mathbf{z}}^-, \bar{\mathbf{z}}^+)$. Noting that $\nabla_{\mathbf{z}^-} J = -\nabla_{\mathbf{z}^+} J = 2\mathbf{M}(\mathbf{z}^- - \mathbf{z}^+)$ it suffices to obtain a contradiction to the stationarity of $(\bar{\mathbf{z}}^-, \bar{\mathbf{z}}^+)$ by $\langle \nabla J, (\Delta\mathbf{z}^-, \Delta\mathbf{z}^+)\rangle < 0$. Inserting the definitions and using $\bar{\mathbf{z}}^+ - \bar{\mathbf{z}}^- = \bar{\mathbf{z}}^+ - \mathbf{z}^+ + \mathbf{z}^+ - \mathbf{z}^- + \mathbf{z}^- - \bar{\mathbf{z}}^-$ we have to show the negativity of

$$\begin{aligned}
&(\bar{\mathbf{z}}^+ - \bar{\mathbf{z}}^-)^T \mathbf{M}(\mathbf{z}^+ - \bar{\mathbf{z}}^+ - \mathbf{z}^- + \bar{\mathbf{z}}^-) \\
&= \quad (\bar{\mathbf{z}}^+ - \mathbf{z}^+)^T \mathbf{M}(-\bar{\mathbf{z}}^+ + \mathbf{z}^+) + (\bar{\mathbf{z}}^+ - \mathbf{z}^+)^T \mathbf{M}(-\mathbf{z}^- + \bar{\mathbf{z}}^-) \\
&\quad +(\mathbf{z}^+ - \mathbf{z}^-)^T \mathbf{M}(\mathbf{z}^+ - \bar{\mathbf{z}}^+ - \mathbf{z}^- + \bar{\mathbf{z}}^-) \\
&\quad +(\mathbf{z}^- - \bar{\mathbf{z}}^-)^T \mathbf{M}(-\mathbf{z}^- + \bar{\mathbf{z}}^-) + (\mathbf{z}^- - \bar{\mathbf{z}}^-)^T \mathbf{M}(-\bar{\mathbf{z}}^+ + \mathbf{z}^+) \\
&=: T_1 + T_2 + T_3 + T_4 + T_5.
\end{aligned}$$

Due to the assumption of positive distance of $\mathbf{z}^+$ and $\mathbf{z}^-$ to the remaining points of their corresponding convex hulls, we have $T_1 < 0$ and $T_4 < 0$. Stationarity of J in $(\mathbf{z}^-, \mathbf{z}^+)$ implies $T_3 \leq 0$.

So it remains to show $T_2 = T_5 \leq 0$, i.e. $(\bar{\mathbf{z}}^+ - \mathbf{z}^+)^T \mathbf{M}(\bar{\mathbf{z}}^- - \mathbf{z}^-) \leq 0$. This follows with the assumption of working in the space $\mathbb{R}^{(1,q)}$. In these spaces the cone of points with positive squared norm and positive first coordinate is closed under additions, positive scalings and reflection with $\mathbf{M}$. Additionally, inner products between such points are always positive. So it remains to show that either both $\bar{\mathbf{z}}^+ - \mathbf{z}^+$ and $\mathbf{z}^- - \bar{\mathbf{z}}^-$ or their negatives lie within this cone. As their squared norm is positive by assumption, it remains to show that they have the same sign in their first component. This can be obtained as both signs must be equal to the sign of the first component of $\mathbf{z}^+ - \mathbf{z}^-$. If e.g. $\bar{\mathbf{z}}^+ - \mathbf{z}^+$ would have different sign in its first component, then the inner product with $\mathbf{z}^+ - \mathbf{z}^-$ would be negative, which would be a contradiction to the stationarity of $(\mathbf{z}^-, \mathbf{z}^+)$. $\square$