A note on the use of warping to impose Dirichlet conditions on Gaussian Processes

Victor Picheny

March 23, 2017

1 Motivating problem

We consider an optimization problem of the type $\min y(\mathbf{x})$ such that $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^d$. We assume that $\mathbb{X}$ is a hyperrectangle.

For some values of a subset of parameters, it is known that y is insensitive to another subset of parameters. Take for instance a function that depends on two discs, parameterized by $x_1 = r_1 \in [0, r_{\max}]$ (radius of the first disc) and $x_2 = \rho_{12} \in [0, 1]$ (ratio between r_1 and the radius of the second disc, r_2). Setting $x_1 = 0$, we have $r_2 = 0$ for any value of ρ_{12}, so $y(0, x_2)$ is constant.

We can define different conditions:

- SIMPLE: if $x_i = c_i$ then y is invariant w.r.t. $\mathbf{x}_J$ (J a subset of $\{1, \dots, d\}$);
- MULTIPLE: if $x_i \in \{c_{i1}, \dots, c_{in}\}$, then y is invariant w.r.t. $\mathbf{x}_J$ (J a subset of $\{1, \dots, d\}$);
- AND: if all $\mathbf{x}_I = \mathbf{c}_I$ then y is invariant w.r.t. $\mathbf{x}_J$ (both I and J are subsets of $\{1, \dots, d\}$, and $I \cap J = \emptyset$);
- OR: if at least one $\mathbf{x}_I = \mathbf{c}_I$ then y is invariant w.r.t. $\mathbf{x}_J$ (both I and J are subsets of $\{1, \dots, d\}$, and $I \cap J = \emptyset$);
- COMBINATION: of the above.

2 Single invariance condition

Here, we consider only a single set I of critical parameters and a single set J of impacted parameters.

2.1 Simple invariance

We first consider the case of a single invariance, $x_i = c_i$, for which a subset $\mathbf{x}_J$ becomes non influent. In the following we use the notation $\mathbf{x} = (x_i, \mathbf{x}_J, \mathbf{x}_{-iJ})$ to make the $\mathbb{X}$ space decomposition explicit (note that this is only notation, no actual permutation is done).

A simple way to handle this problem is to distort locally the space so that the subspace $\{(x_i, \mathbf{x}_J) | x_i = c_i\}$ collapses to a single point, for instance with $\mathbf{x}_J$ at its average value: $(c_i, \overline{\mathbf{x}_J})$. Hence, we are seeking warping functions of the form:

$$\begin{aligned} \psi : \mathbb{X} &\rightarrow \widetilde{\mathbb{X}} \\ \mathbf{x} &\mapsto \widetilde{\mathbf{x}} \end{aligned}$$

such that:

1. $\psi(x_i, \mathbf{x}_J, \mathbf{x}_{-iJ}) = (c_i, \overline{\mathbf{x}_J}, \mathbf{x}_{-iJ})$ if and only if $x_i = c_i$;
2. ψ restricted to $\mathbb{X} \backslash (c_i, ., .)$ and $\widetilde{\mathbb{X}} \backslash (c_i, \overline{\mathbf{x}_J}, .)$ is a diffeomorphism;

3. the deformation decreases monotonically when $|x_i - c_i|$ increases, that is:

$$d\left(\left(x_i^k, \mathbf{x}_J, \mathbf{x}_{-iJ}\right), \psi\left[\left(x_i^k, \mathbf{x}_J, \mathbf{x}_{-iJ}\right)\right]\right) \leq d\left(\left(x_i^l, \mathbf{x}_J, \mathbf{x}_{-iJ}\right), \psi\left[\left(x_i^l, \mathbf{x}_J, \mathbf{x}_{-iJ}\right)\right]\right) \text{ if } |x_i^k - c_i| \leq |x_i^l - c_i|$$

for some distance $d(.,.)$.

Assuming that the $\mathbf{x}_J$ dimension collapses to $\overline{\mathbf{x}_J}$ at $x_i = c_i$, we write:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j})\, \alpha(x_i, c_i), \tag{1}$$

with $\alpha(x_i, c_i)$ an attenuation function such that:

1. $\alpha(c_i, c_i) = 0$;
2. α increases monotonically with $|x_i - c_i|$;
3. $0 < \alpha \leq 1$, $\forall x_i \neq c_i$.

Condition 1 ensures that $\widetilde{x_j} = \overline{x_j}$ when $x_i = c_i$ (the j-th dimension collapses).

We propose linear and correlation-based attenuation functions:

$$\alpha_{\text{lin}}(x_i, c_i) = \frac{|x_i - c_i|}{\delta_i}, \tag{2}$$

$$\alpha_{\text{cor}}(x_i, c_i) = 1 - r(x_i, c_i), \tag{3}$$

where r is a $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ correlation function. Typically, δ_i may be set to the range of variation of x_i, so that the condition $\alpha \leq 1$ is ensured. Choosing r as the generalized exponential correlation, we have:

$$\alpha_{\text{exp}}(x_i, c_i) = 1 - \exp\left[-\left(\frac{|x_i - c_i|}{\theta_i}\right)^d\right], \tag{4}$$

with θ_i and d positive parameters to be tuned.

Figure 1 shows a 2D rectangular space distorted by three warpings, when the invariance is on a boundary of x_1. Figure 2 shows (unconditional) realizations of GPs with a Gaussian kernel applied on the warped space. We see that the invariance at x_1 maximum is ensured. The linear warping induces a strong anisotropy, while with the two other warpings, the process seems stationary far from the critical value.

2.2 "AND" invariance

Now, we consider the case where invariances occur when a set of variables takes simultaneously a set of critical values: $\mathbf{x}_I = \mathbf{c}_I$. On a cubic space, this amounts to imposing invariance on an edge[1]. In that case, a possible warping is:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j})\, \alpha_I(\mathbf{x}_I, \mathbf{c}_I). \tag{5}$$

with α_I now a multivariate attenuation function ($\mathbb{R}^{\text{Card}(I)} \times \mathbb{R}^{\text{Card}(I)} \to \mathbb{R}$), so that, similarly to the simple case:

1. $\alpha_I(\mathbf{c}_I, \mathbf{c}_I) = 0$;
2. α_I increases monotonically with $d(\mathbf{x}_I, \mathbf{c}_I)$ (for some distance $d(.,.)$;
3. $0 < \alpha_I \leq 1$, $\forall \mathbf{x}_I \neq \mathbf{c}_I$.

As in the simple case, linear and correlation-based warpings can be defined as:

$$\alpha_{\text{lin}}(\mathbf{x}_I, \mathbf{c}_I) = \frac{1}{\text{Card}(I)} \sum_{i \in I} \frac{|x_i - c_i|}{\delta_i}, \tag{6}$$

$$\alpha_{\text{cor}}(\mathbf{x}_I, \mathbf{c}_I) = 1 - r_I(\mathbf{x}_I, \mathbf{c}_I), \tag{7}$$

[1]such a condition cannot exist in 2D under the assumption $I \cap J = \emptyset$, since $\text{Card}(J) \geq 2$.

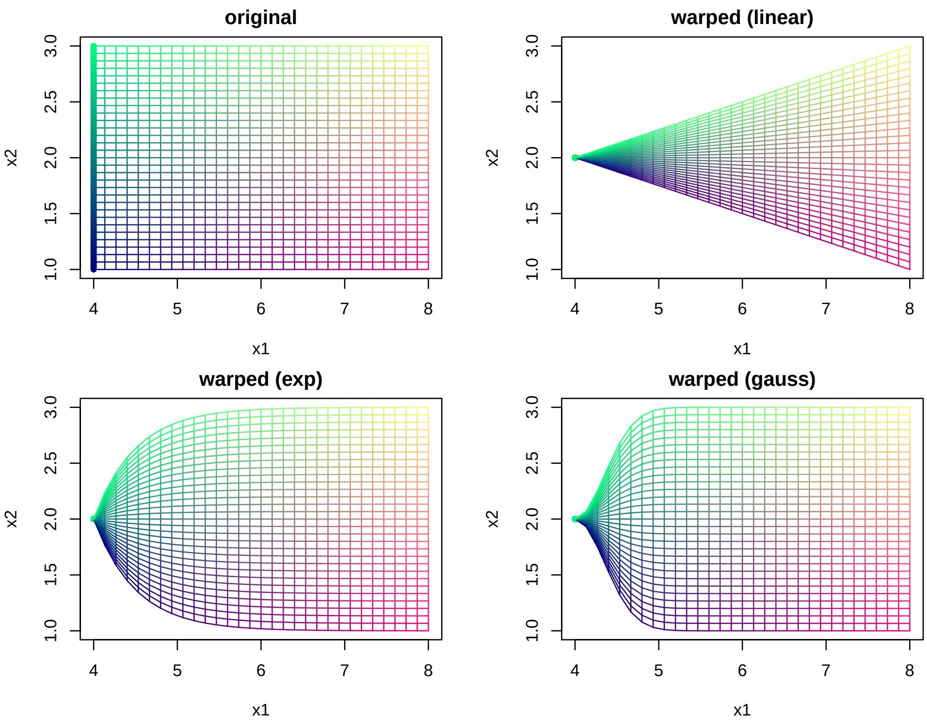

Figure 1: Three deformations of a 2D space. The local invariance is at $x_1 = 0$, highlighted with larger lines.

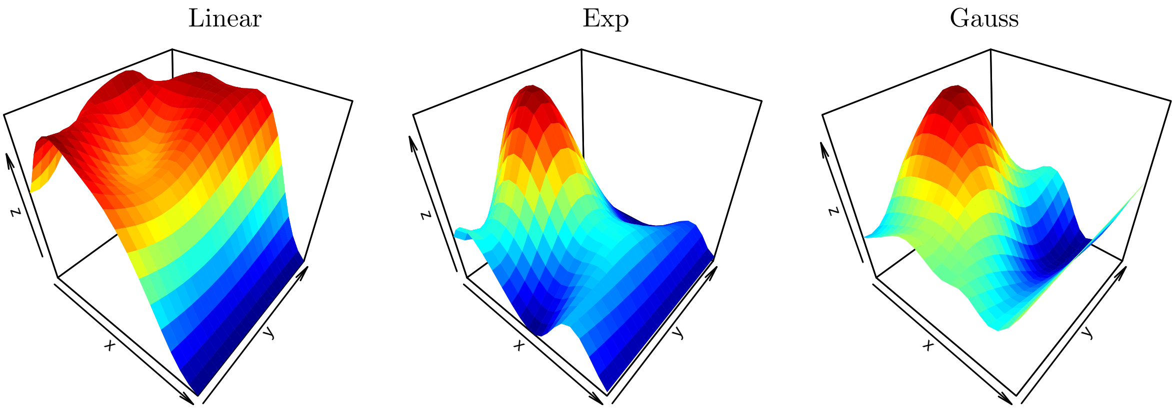

Figure 2: Three GP realizations using warping functions as shown previously.

with r_I a $\mathbb{R}^{\text{Card}(I)} \times \mathbb{R}^{\text{Card}(I)} \to \mathbb{R}$ correlation function, for instance:

$$r_I(\mathbf{x}_I, \mathbf{c}_I) = \exp\left[-\sum_{i \in I} \left(\frac{|x_i - c_i|}{\theta_i}\right)^d\right]$$

Figure 3 shows a deformation of a cubic space when x_3 is not influent when both x_1 and x_2 are minimal, when a Gaussian warping (exponential with $d = 2$) is applied.

Affine conditions The "AND" condition can be generalized to the following affine formulation: y is invariant w.r.t. $\mathbf{x}_J$ if $\mathbf{A}\mathbf{x}_I = \mathbf{b}$, with $\mathbf{A}$ a matrix of size $p \times \text{Card}(I)$ and $\mathbf{b}$ a vector of size p. For the "AND" condition, we have $\mathbf{A} = \mathbb{I}_p$ and $\mathbf{b} = \mathbf{c}_I$. The warping function is the same as in Equation 5, with now:

$$\alpha(\mathbf{x}_I, \mathbf{c}_I) \quad = \quad 1 - r_A(\mathbf{A}\mathbf{x}_I, \mathbf{b}). \tag{8}$$

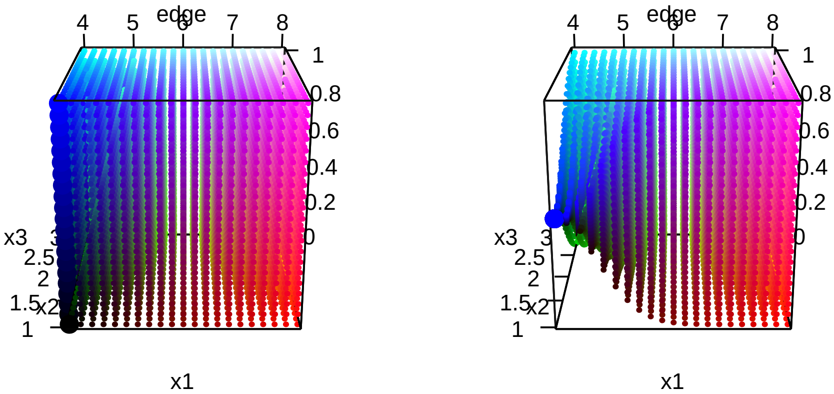

Figure 3: Warping for $c_1 = R$ AND $c_2 = 1$. Left: original space (with the critical edge highlighted), right: distorted space.

Note that choosing the range of the correlation r_A is non-trivial. A possible solution is $\theta_A = \mathbf{A}^T \boldsymbol{\theta}_I$.

Figure 4 show the deformation of a cubic space with the condition: y invariant w.r.t. x_3 if $x_1 = x_2$.

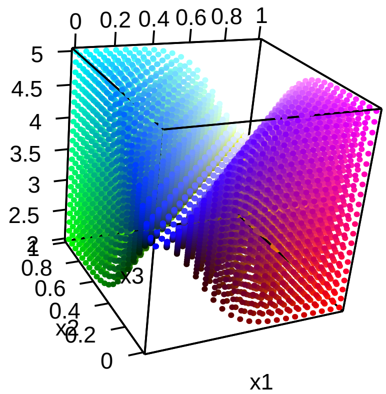

Figure 4: Warping for the affine condition: y invariant w.r.t. x_3 if $x_1 = x_2$.

3 Combining warpings

3.1 Independent conditions

Now, we consider that we have a series of invariance conditions, defined with respect to sets $I_1, \ldots, I_n$ and corresponding $J_1, \ldots, J_n$. If $J_k \cap J_l = \emptyset$, $1 \leq j \neq k \leq n$ and $I_i \cap J_k = \emptyset$, $1 \leq j, k \leq n$, the set of warped variables are distinct from the set on which the conditions are written, the invariance conditions are written only once for each variable. In that case, the warpings can be applied independently.

3.2 Combinations of simple conditions: "OR" invariance

Now, we consider the case when y is invariant w.r.t. a set $\mathbf{x}_J$ for different conditions on sets $I_1, \ldots, I_n$ (that, for $\mathbf{x}_{I_1} = \mathbf{c}_{I_1}$ OR $\mathbf{x}_{I_2} = \mathbf{c}_{I_2}$ OR $\ldots$). If $J \cap I_i = \emptyset$, $1 \leq i \leq n$, the warping function we propose is:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j}) \prod_{I \in \{I_1, \ldots, I_n\}} \alpha_I(\mathbf{x}_I, \mathbf{c}_I). \tag{9}$$

We see directly that the product of α's ensure that $\widetilde{x_j} = \overline{x_j}$ if any $x_i = c_i$, and the distortion reduces only when *all* the x_i's are far from the c_i's. Figure 5 shows a deformation of a cubic space when x_3 is not influent when x_1 or x_2 are minimal, when a Gaussian warping (exponential with $d = 2$) is applied.

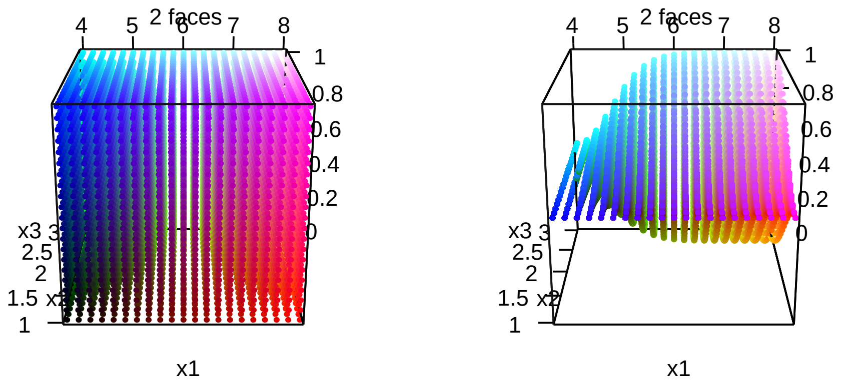

Figure 5: Warping for $c_1 = 4$ (left face of the cube) OR $c_2 = 1$ (front face). Left: original space, right: distorted space.

Multiple conditions on the same variable If y is invariant w.r.t. $\mathbf{x}_J$ for any of the conditions $x_i = \{c_{i1}, \ldots, c_{in}\}$, the warping function is:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j}) \prod_{k=1}^{n} \alpha(x_i, c_{ik}). \tag{10}$$

An example if given in Figure 6, using a Gaussian attenuation.

Note that the c_{ik}'s can take any value and not necessarily a boundary: here we used $c_{11} = 4$ (lower bound) and $c_{12} = 7$ (in the middle of the domain of x_1). On can notice that far away from the critical conditions, the mesh is undeformed locally AND globally, and in particular the distance between the far left and far right of the figure is unchanged by the warping.

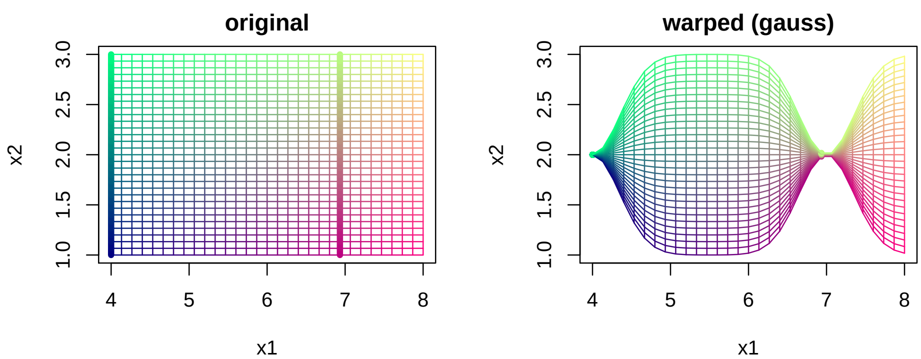

Figure 6: Warping for two critical values for x_1.

3.3 "Circular" conditions

Difficulty only arises when some variables appear in both I_l's and J_m's sets. Take for instance a "reciprocal" condition, e.g., y is invariant w.r.t. $\mathbf{x}_J$ when $\mathbf{x}_I = \mathbf{c}_I$, and invariant w.r.t. $\mathbf{x}_I$ when $\mathbf{x}_J = \mathbf{c}_J$. In that case, applying independently warping functions would lead to:

$$\begin{aligned} \psi(\mathbf{c}_I, \mathbf{x}_J, \mathbf{x}_{-IJ}) &= (\mathbf{c}_I, \overline{\mathbf{x}_J}, \mathbf{x}_{-IJ}), \\ \psi(\mathbf{x}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}) &= (\overline{\mathbf{x}_I}, \mathbf{c}_J, \mathbf{x}_{-IJ}), \\ \text{but: } \psi(\mathbf{c}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}) &= (\mathbf{c}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}), \end{aligned}$$

which induces a discontinuity.

In that case, a simple solution is to fix the non influent variable to its critical value instead of its average, hence applying:

$$\forall k \in K = \left(\cup_{1\leq l \leq n} I_l\right) \cap \left(\cup_{1\leq m \leq n} J_m\right), \quad \widetilde{x_k} = \quad c_k + (x_k - c_k) \prod_{i \in I_k} \alpha(x_i, c_i). \tag{11}$$

Remark This formula does not apply when x_k takes multiple critical values (Equation 10) or in the affine case (Equation 8).

We first show the deformations on a 2D space on Figure 7, where the two critical values are on the boundaries of x_1 and x_2, and on Figure 8, where one critical value is in the middle of the x_1 space. Here, the warping of Equation 11 is applied on each variable ($K = \{1, 2\}$). Again, except for the linear warping, the local topology is preserved far from the critical edges. On Figure 8, we also see that long-range distances are also unchanged. A GP realization is given on Figure 9. The "T-shaped" invariance is ensured, and the GP is stationary far from the critical values.

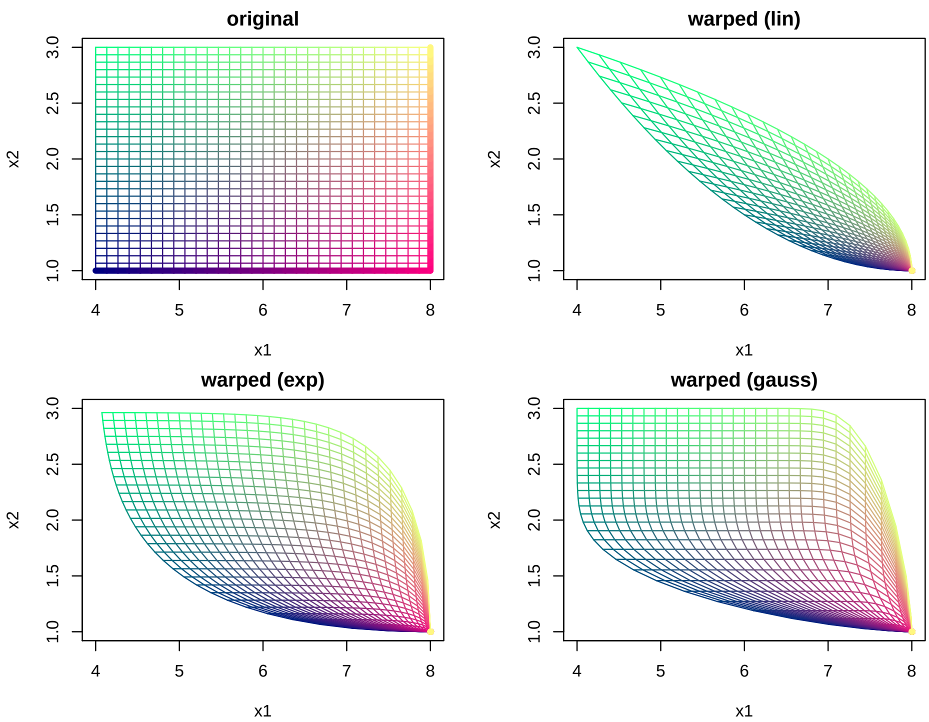

Figure 7: Three deformations of a 2D space, with invariance at $x_1 = 8$ OR $x_2 = 1$, highlighted with larger lines.

Then, we consider a cubic space with the following circular conditions:

- y is invariant w.r.t. x_2 if $x_1 = 4$;
- y is invariant w.r.t. x_3 if $x_2 = 1$;

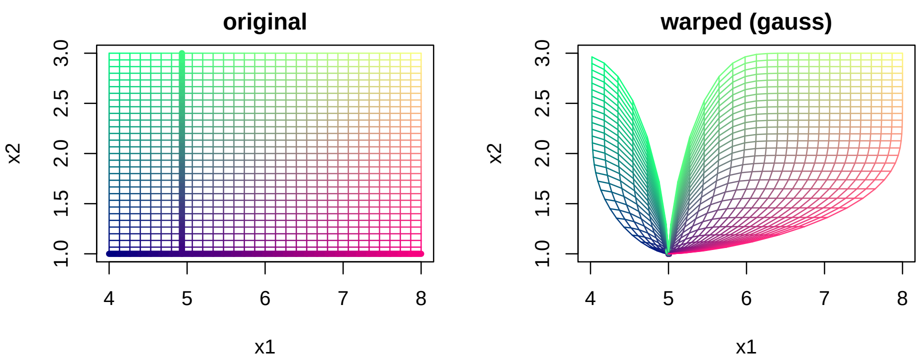

Figure 8: Warping for a T-shaped “OR” invariance.

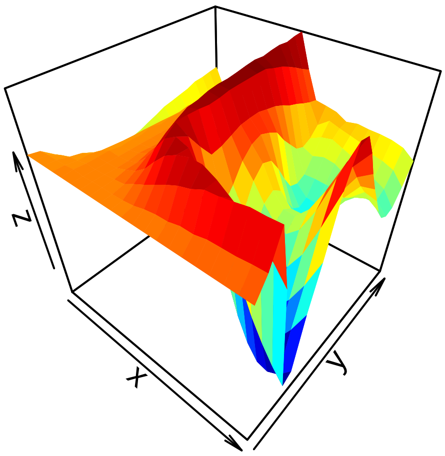

Figure 9: GP realization for a T-shaped “OR” invariance.

- y is invariant w.r.t. x_1 if $x_3 = 0$.

All critical values correspond to the lower bounds of the variables. Equation 11 is applied to each variable, hence with $K = \{1, 2, 3\}$, $C = [4, 1, 0]$ and $I_1 = 3$, $I_2 = 1$, and $I_3 = 2$. The original and distorted space is shown in Figure 10.

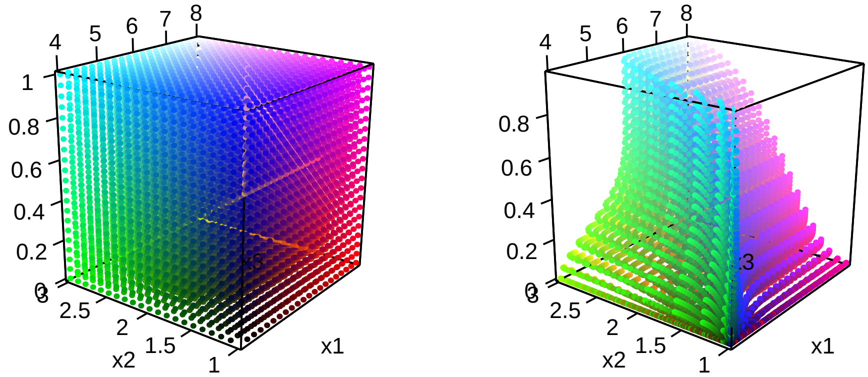

Figure 10: Warping with circular conditions. Left: original space, right: distorted space.

3.4 "Circular" affine conditions

In this section, we report experimental solutions that appear to work on 3D cases. We consider a series of three test cases, on the design space $x_1 \in [1, 4]$, $x_2 \in [2, 3]$ and $x_1 \in [0, 2]$, defined by the following conditions:

- (affine + simple circular conditions) y is invariant w.r.t. x_1 if $2x_2 = 3x_3$, and w.r.t. x_3 if $x_1 = 1$;
- (2 affine circular conditions) y is invariant w.r.t. x_1 if $2x_2 = 3x_3$, and w.r.t. x_2 if $x_1 = 2x_3$;
- (3 affine circular conditinos) y is invariant w.r.t. x_1 if $2x_2 = 3x_3$, w.r.t. x_3 if $x_1 = 1$ and w.r.t. x_3 if $3x_1 = 4x_2$.

Note that the design space is voluntarily different from the unit cube.

In the first case, we applied:

$$\begin{aligned}
\widetilde{x_1} &= c_1 + (x_1 - c_1)\,\alpha(x_1, c_1)\alpha(2x_2 - 3x_3, 0) \\
\widetilde{x_3} &= c_3 + (x_3 - c_3)\,\alpha(x_i, c_1)\alpha(2x_2 - 3x_3, 0).
\end{aligned}$$

In the second case, we applied:

$$\begin{aligned}
\widetilde{x_1} &= c_1 + (x_1 - c_1)\,\alpha(x_1, c_1)\alpha(2x_2 - 3x_3, 0) \\
\widetilde{x_2} &= c_2 + (x_1 - c_2)\,\alpha(x_1, c_1)\alpha(2x_2 - 3x_3, 0) \\
\widetilde{x_3} &= c_3 + (x_3 - c_3)\,\alpha(x_i, c_1)\alpha(2x_2 - 3x_3, 0).
\end{aligned}$$

In the third case, we applied:

$$\begin{aligned}
\widetilde{x_1} &= c_1 + (x_1 - c_1)\,\alpha(x_1, c_1)\alpha(2x_2 - 3x_3, 0)\alpha(3x_1 - 4x_2, 0) \\
\widetilde{x_2} &= c_2 + (x_1 - c_2)\,\alpha(x_1, c_1)\alpha(2x_2 - 3x_3, 0)\alpha(3x_1 - 4x_2, 0) \\
\widetilde{x_3} &= c_3 + (x_3 - c_3)\,\alpha(x_i, c_1)\alpha(2x_2 - 3x_3, 0)\alpha(3x_1 - 4x_2, 0).
\end{aligned}$$

On all cases, C was chosen within the intersection of the critical plans: for instance in the second example, at the intersection of $2x_2 = 3x_3$, and $x_1 = 2x_3$. Since this intersection is a segment, we took C as the the average of the intersection.

Figure 11

3.5 A complex example

Finally, we show the deformations on a cubic space, with the following conditions:

- y is invariant w.r.t. x_3 if $x_1 = 4$ or $x_2 = 1$
- y is invariant w.r.t. x_2 if $x_1 = 4$
- y is invariant w.r.t. x_1 if $x_2 = 1$

Equation 9 is applied to x_3 ($j = 3$, $I = \{1, 2\}$), while Equation 11 is applied to x_1 and x_2 ($K = \{1, 2\}$, $I_1 = 2$ and $I_2 = 1$), with a Gaussian warping. Invariance occurs on the left and front faces of the cube (x_1 and x_2 to their minimum, Figure 12, left). In the resulting topology, those two faces collapse into a single point as desired. Note however the large difference with the "AND" condition in Figure 3, for which the cubic topology is mostly preserved except close to the critical edge.

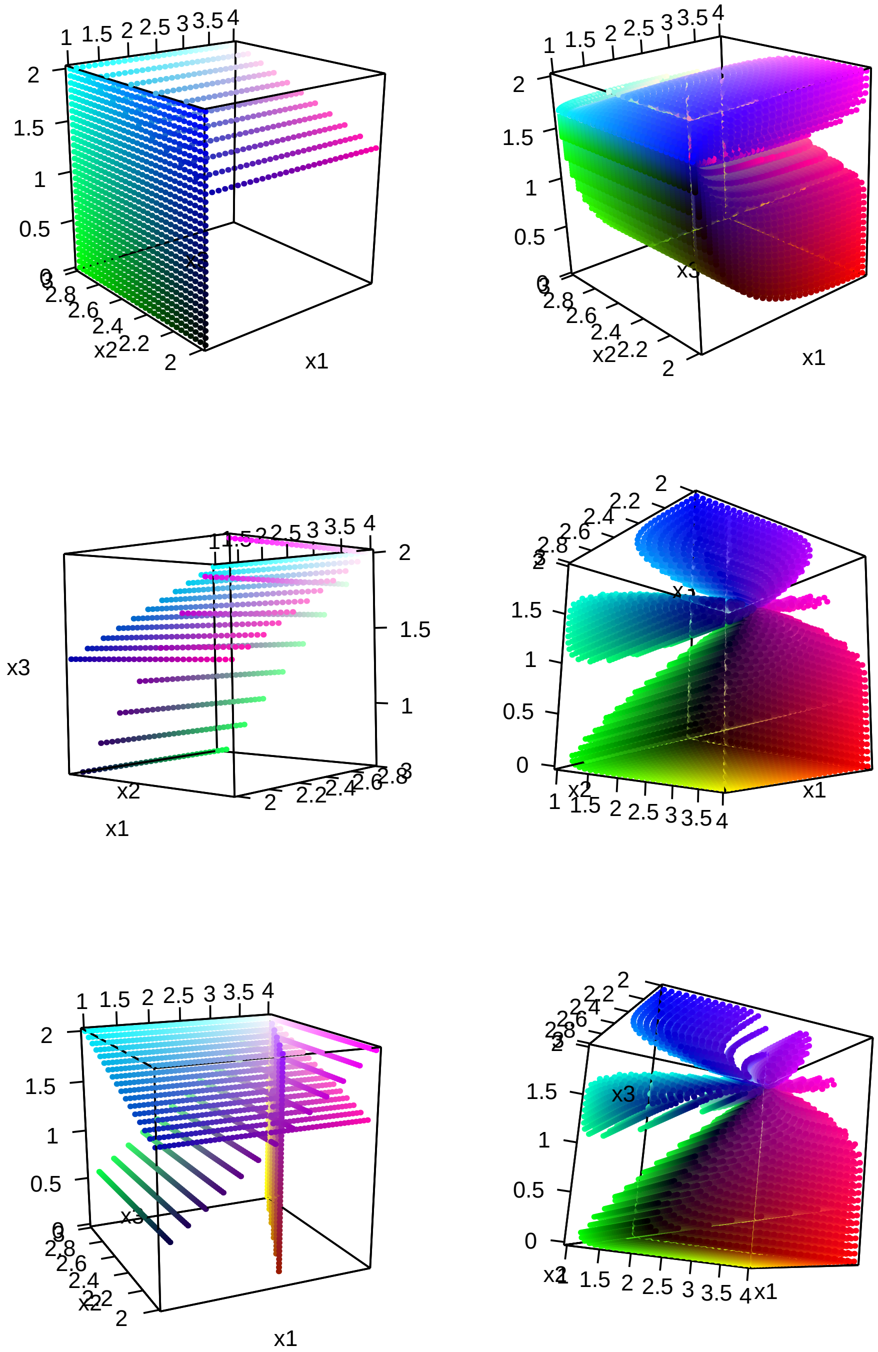

Figure 11: Warpings for three different cases with circular conditions. Left: critical plans, right: corresponding distorted space.

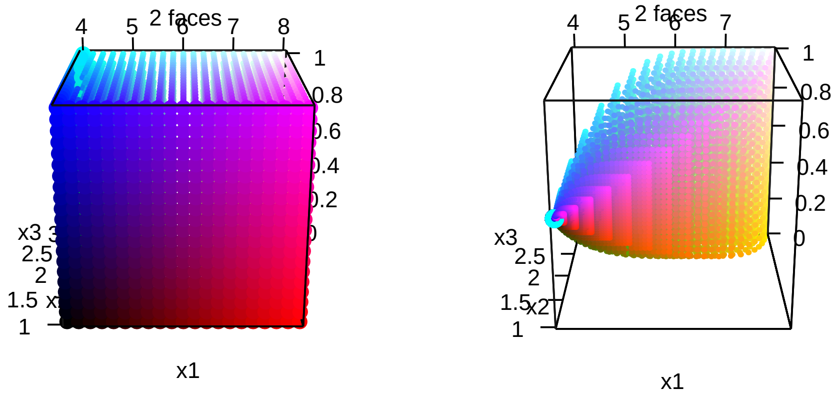

Figure 12: Warping for x_2 and x_3 inactive if $c_1 = 4$ and x_1 and x_3 inactive if $c_2 = 1$. Left: original space, right: distorted space.