

Document downloaded from:

<http://hdl.handle.net/10251/169648>

This paper must be cited as:

Palací-López, D.; Villalba-Torán, PM.; Facco, P.; Barolo, M.; Ferrer, A. (2020). Improved formulation of the latent variable model inversion; based optimization problem for quality by design applications. *Journal of Chemometrics*. 34(6):1-18. <https://doi.org/10.1002/CEM.3230>



The final publication is available at

<https://doi.org/10.1002/CEM.3230>

Copyright John Wiley & Sons

Additional Information

Table of contents abstract

Improved formulation of the Latent Variable Model Inversion-based optimization problem for Quality by Design applications

Daniel Palací-López, Pedro Villalba, Pierantonio Facco, Massimiliano Barolo, Alberto Ferrer

A new formulation of the PLS-regression model inversion-based optimization problem is proposed that allows, using happenstance data: i) optimizing quality attributes defined as linear combinations of variables; ii) sequentially obtaining new solutions without having to actively modify the constraints imposed on them, iii) avoiding the need to define arbitrarily low/high values as desired ones for each quality attribute for their minimization/maximization, and iv) obtaining a wider range of sets of inputs that guarantee the desired values for the quality attributes.

**Improved formulation of the Latent Variable Model Inversion-based optimization
problem for Quality by Design applications**

Palací-López, D.^{a,*}, Villalba, P^b. Facco, P.^{c,**}, Barolo, M.^c, Ferrer, A.^a

a) *Multivariate Statistical Engineering Group, Department of Applied Statistics,
Operational Research and Quality, Universitat Politècnica de València, Camino de
Vera s/n, 7A, 46022, Valencia, Spain*

b) *Instituto Tecnológico de Informática, Ciudad Politécnica de la Innovación,
Universitat Politècnica de València, Camino de Vera s/n, 8G, Acceso B, 46022,
Valencia, Spain*

c) *CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of
Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova PD, Italy*

* Corresponding author: Daniel Palací-López. Departamento de Estadística e
Investigación Operativa Aplicadas y Calidad. Universitat Politècnica de València. Cno.
De Vera s/n, Edificio 7A, 46022, Valencia, Spain

E-mail: dapalpe@gmail.com

** Corresponding author: Pierantonio Facco. Computer-Aided Process Engineering
Laboratory, Department of Industrial Engineering, University of Padova, phone:
+39.049.827.5470. Via Marzolo 9, 35131 Padova PD, Italy.

E-mail: pierantonio.facco@unipd.it

Abstract

Latent Variable Regression Model (LVRM) inversion is a relevant tool for finding, if they exist, different combinations of manufacturing conditions that yield the desired process outputs. Finding the best manufacturing conditions can be done by optimizing an appropriately formulated objective function using nonlinear programming. To this end, different formulations of the optimization problem based on LVRM inversion have been proposed in the literature that allow the use of happenstance data (e.g. historical data) for this purpose, present lower computational costs than optimizing in the space of the original variables, and guarantee that the solution will conform to the correlation structure of available data from the past. However, these approaches, as presented, suffer from some limitations, such as having to actively modify the constraints imposed on the solution to achieve different sets of conditions to those available in the LVRM calibration dataset, or the lack of a standardized approach for optimizing a linear combination of variables. Furthermore, when minimizing or maximizing one or more outputs, a severe handicap is also present related to the definition of arbitrarily low or high ‘desired’ values. This paper aims at tackling all of these issues. The resulting proposed formulation of the optimization problem is illustrated with three case studies.

Keywords: Partial Least-Squares (PLS); Latent Variable modelling; Latent Variable model inversion; Optimization in the latent space; Quality by design (QbD)

1. Introduction

The Quality-by-Design (QbD)^{1,2} initiative promotes the implementation of science-based methodologies to deliver products that meet the desired specifications while increasing the flexibility and robustness of the processes. This is done by allowing changes in materials and processing conditions without negatively affecting the outputs’

quality. In this context, process optimization can be understood as the identification, through whatever means, of different sets of conditions that belong to the so-called design space (DS), defined as “the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality”³.

Model-based process optimization requires building a causal model that relates changes in the process inputs (e.g. raw materials properties, processing conditions, etc.) with those in the process outputs (e.g. critical quality attributes, productivity, profitability, etc.). Although deterministic models based on first principles are always desirable for this purpose, insufficient knowledge and/or resources required to properly construct them makes their use infeasible in practice in most cases. Empirical, i.e. data-driven, models are often built instead^{4,5}. Nevertheless, in order to use them for optimization, the data used to construct them must be appropriate to guarantee causality⁶. When classical linear regression or machine learning models are resorted to, independent variation in the inputs is required. This implies that data from design of experiments (DOE)⁷ must be gathered. Due to the usually large amount of inputs and outputs involved this requires extensive experimentation that is infeasible in practice.

Alternative methods based on latent variables (LV), such as partial least squares (PLS) regression-based techniques, allow analysing large datasets with highly correlated and low signal-to-noise ratio data. This makes them especially useful to deal with big data in the implementation of the Industry 4.0 paradigm, when happenstance data (e.g. “routine” data coming not from a DOE but from daily production, and therefore not causal in nature) are collected. PLS not only models the relationship between the input (X) space and the output (Y) space, but also provides models for both spaces. This gives

PLS models two powerful properties: uniqueness and causality in the latent space no matter if data come from either DOE or daily production⁸⁻¹⁰. Therefore, it can be used in process optimization to find the combinations of input variables that are consistent with their past correlation structure and also guarantee the desired output values.

In order to apply the QbD initiative, two distinct strategies have been proposed in the literature: i) defining or estimating the DS as a whole, and; ii) solving an optimization problem in an attempt to obtain single sets of process conditions within the DS.

The first of these strategies, when resorting to LV-based methods, relies on the so-called null space (NS), i.e. the subspace in the latent space within which the prediction of the outcome variables \hat{y} does not vary^{11,12}, for which the uncertainty in its definition can also be accounted for¹³⁻¹⁶. Due to the nature of data-driven approaches, it is unadvisable to extrapolate any results provided by the model outside of the so-called knowledge space (KS), i.e. the regions constituted by the set of historical products that have already been manufactured^{17,18}. Additional constraints on e.g. process operating conditions may also be accounted for in order to further delimit a portion of the DS inside the KS within which products of the desired quality may still be produced while meeting such constraints¹⁶.

As for the second strategy, solving an optimization problem where the same constraints are considered allows obtaining individual sets of operating conditions that guarantee the desired outputs (i.e., samples expected to be within the DS, inside the KS, and meet all other restrictions). Such problem may be formulated in the space of the original variables and imposing an upper limit on the acceptable Euclidean distance to the latent subspace of a previously fitted latent variable regression model (LVRM), so that any solution (i.e., values for the input variables) obtained abides, to an extent (but not

strictly), by the correlation structure found by such LVRMⁱ. This formulation, however, suffers from high computational cost and may be even infeasible in practice as the number of input and output variables increases. Accordingly, an optimization problem in the space of the latent variables can be formulated, significantly reducing the computational cost. However, such approach provides solutions that necessarily abide by the correlation structure found by the LVRM, therefore losing the flexibility of accepting small changes on the correlation structure among variables. This limitation, already discussed in the literature¹⁹, can be solved with a two-step optimization process²⁰, where the optimization problem is solved in the latent space as a first step, and then the solution is used as a starting point for the optimization problem in the original space, providing some degrees of freedom to obtain more flexible solutions.

The formulation of the aforementioned approach, however, still presents some limitations that make it unfit in some scenarios:

- i) when a quality attribute of interest can or should be expressed as a linear combination of variables, no standardized approach to optimize it has yet been proposed. This may happen, for example, when the quality attribute of interest itself (e.g. concentration of the main component in a blend product) is poorly explained by the model, but this is not the case for the variables from which it can be calculated (e.g. amount of the main component and total amount of the resulting blend)¹⁶. As an additional example, suppose that the (quality) attribute of interest is the income generated by selling

ⁱ Note that this is not the same as optimizing via e.g. Machine Learning techniques in the space of the original variables and without resorting to LV-based techniques. In such case, causality cannot be inferred and there is no guarantee that the solution would respect the correlation structure of the data used to fit the model, leading to infeasible solutions

several products. This income may be expressed as a linear combination between the amount of each single products and the relevant selling price;

- ii) if new solutions in successive optimizations, or solutions different from already available observations are desired, current formulations require actively modifying either the hard constraints imposed on the solution or the weight given to each term in the objective function;
- iii) given the way the objective function is usually formulated, ‘desired values’ for the variables included in the objective function must be defined, such that the value of the objective function will be optimized if/when such values are achieved for the solution. However, no standardized approach has yet been offered to define such ‘desired values’ when the goal of the optimization problem is to minimize/maximize them, especially if no feasible or known minimum/maximum values are known.

This paper aims at providing an alternative, more general formulation of the optimization problem in the space of the latent variables that also accounts for these three issues. Section 2 introduces one of the most common formulations of the optimization problem in the latent space as proposed in the literature. This will be used as a basis for the new proposed formulation. Section 3 is divided in three different subsections where the optimization problem is subsequently reformulated to tackle the three aforementioned limitations and illustrates the usefulness of the proposal with three different case studies. Conclusions are presented in Section 4.

2. Optimization in the latent space

Partial Least Squares (PLS) regression²¹ is among the most used latent variable-based methods to build data-based models for optimization. It models the inner relationships

between the matrices of inputs \mathbf{X} [$N \times M$] and outputs \mathbf{Y} [$N \times L$] in order to predict \mathbf{Y} from the A -dimensional subspace of the latent variables associated to \mathbf{X} , such that its covariance with \mathbf{Y} is maximised. This model structure can be expressed as²¹:

$$\begin{aligned}\mathbf{X} &= \mathbf{T} \cdot \mathbf{P}^T \cdot \mathbf{D}_{s_x} + \mathbf{1} \cdot \mathbf{m}_X^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{T} \cdot \mathbf{Q}^T \cdot \mathbf{D}_{s_y} + \mathbf{1} \cdot \mathbf{m}_Y^T + \mathbf{F} \\ \mathbf{T} &= (\mathbf{X} - \mathbf{1} \cdot \mathbf{m}_X^T) \cdot \mathbf{D}_{s_x}^{-1} \cdot \mathbf{W} \cdot (\mathbf{P}^T \cdot \mathbf{W})^{-1} = (\mathbf{X} - \mathbf{1} \cdot \mathbf{m}_X^T) \cdot \mathbf{D}_{s_x}^{-1} \cdot \mathbf{W}^*\end{aligned}\tag{1}$$

being \mathbf{T} [$N \times A$], \mathbf{P} [$M \times A$] and \mathbf{E} [$N \times M$] the \mathbf{X} scores, loadings and residuals matrices, respectively; \mathbf{Q} [$L \times A$] and \mathbf{F} [$N \times L$] the respective \mathbf{Y} loadings and residuals matrix, and \mathbf{W} [$M \times A$] the weighting matrix, such that $A \leq r_x$, r_x being the rank of \mathbf{X} ; \mathbf{m}_X and \mathbf{m}_Y the [$M \times 1$] and [$L \times 1$] column vectors of centring factors for the inputs and outputs, respectively; \mathbf{D}_{s_x} and \mathbf{D}_{s_y} the [$M \times M$] and [$L \times L$] diagonal matrices with the scaling factors applied to the input and output variables, respectively, before fitting the PLS-regression model; and $\mathbf{1}$ a vector of ones.

If at least one feasible combination of inputs exists that leads to the desired outputs, Latent Variable Regression Model Inversion (LVRMI) can be used to obtain a combination of inputs that, according to the model, guarantees the desired outputs. If more than one combination exists, the so-called Null Space (NS), i.e. the projection onto the latent space of all the combinations of inputs theoretically guarantying the desired outputs, can also be defined^{11,12}. This definition of the NS was also extended to quality attributes defined as linear combinations of other variables¹⁶.

Alternatively, or in order to find the best feasible combination of inputs, an optimization problem in the latent variables space can be resorted to¹³, formulated as follows:

$$\begin{aligned}
& \min_{\boldsymbol{\tau}} \left[g_0 \cdot (\hat{\mathbf{y}}_{\text{NEW}} - \mathbf{y}_{\text{DES}})^T \cdot \boldsymbol{\Gamma} \cdot \mathbf{D}_{\text{SY}}^{-2} \cdot (\hat{\mathbf{y}}_{\text{NEW}} - \mathbf{y}_{\text{DES}}) + g_1 \cdot \sum_{a=1}^A \frac{\tau_a^2}{s_a^2} \right] \\
& \text{s. t.} \\
& \hat{\mathbf{y}}_{\text{NEW}} = \mathbf{D}_{\text{SY}} \cdot \mathbf{Q} \cdot \boldsymbol{\tau} + \mathbf{m}_{\text{Y}} \\
& \hat{\mathbf{x}}_{\text{NEW}} = \mathbf{D}_{\text{SX}} \cdot \mathbf{P} \cdot \boldsymbol{\tau} + \mathbf{m}_{\text{X}} \\
& T_{\boldsymbol{\tau}}^2 = \boldsymbol{\tau}^T \cdot \boldsymbol{\Lambda}^{-1} \cdot \boldsymbol{\tau} \leq T_{\text{lim}}^2 \\
& \mathbf{A}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} \leq \mathbf{d}_{\boldsymbol{\tau}} \\
& \mathbf{F}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} = \mathbf{f}_{\boldsymbol{\tau}}
\end{aligned} \tag{2}$$

Where $\boldsymbol{\tau}$ is the score vector solution, $\boldsymbol{\Gamma}$ is a $[L \times L]$ diagonal matrix where the l -th element in the diagonal represents the weight given to achieving the desired value for the l -th output variable (such that the sum of the absolute values of the diagonal elements is one); g_0 and g_1 are the weights given to each term in the objective function when solving the optimization problem; $\hat{\mathbf{x}}_{\text{NEW}}$ and $\hat{\mathbf{y}}_{\text{NEW}}$ are the vectors of inputs and outputs, respectively, predicted by the PLS model given $\boldsymbol{\tau}$; $\boldsymbol{\Lambda}^{-1}$ is the $[A \times A]$ diagonal matrix containing the inverse of the A variances of the scores (s_a^2) associated to the LVs; $\mathbf{A}_{\boldsymbol{\tau}}$ and $\mathbf{d}_{\boldsymbol{\tau}}$ (and $\mathbf{F}_{\boldsymbol{\tau}}$ and $\mathbf{f}_{\boldsymbol{\tau}}$) are a matrix and a vector used to define inequality (and equality) hard constraints on the latent variables, respectivelyⁱⁱ. Confidence limits T_{lim}^2 and SPE_{lim} can be accounted for²² but, in this case, since any $\hat{\mathbf{x}}_{\text{NEW}}$ obtained from $\boldsymbol{\tau}$ will fall exactly within the model subspace, its SPE will be zero, i.e. $SPE_{\hat{\mathbf{x}}_{\text{NEW}}} = 0$. Therefore, no constraint on $SPE_{\hat{\mathbf{x}}_{\text{NEW}}}$ needs to be defined.

Note that, in this formulation, two different kinds of constraints are considered. ‘Hard constraints’, defined outside of the objective function, are restrictions that determine the envelope of the subspace of acceptable solutions for the optimization problem, but do not otherwise favour any potential solution over the rest. On the other hand, terms in the objective function are ‘soft constraints’, as they favour solutions being close to (or far

ⁱⁱ Restriction on the inputs/outputs can also be transferred to the latent space to be expressed as restrictions on the latent variables¹⁶

from) specific regions inside this subspace, but do not affect the shape or size of the region of acceptable solutions.

This formulation presents a series of limitations:

1. It does not allow the proper optimization of any quality attribute, d_r , that can be expressed as a linear combination of outputs, i.e. $d_r = \mathbf{a}_r^T \cdot \mathbf{y}$. This is because the soft constraints in the objective function associated to outputs (\mathbf{Y}) and/or inputs (\mathbf{X}) are imposed on each individual variable (i.e. not linear combinations of them). While modifying the weight given to the soft constraints associated to achieving the desired values for these variables may seem like a proper approach to bypass this issue, however, such is not the case (this is explained in more depth in Section 3.1 and demonstrated in the Appendix).
2. Obtaining new solutions in subsequent optimizations and/or solutions that span previously unexplored areas of the knowledge space will rely on changes in the hard constraints imposed on these solutions (on the inputs, outputs, or any combination of them) and/or in the weight given to each term in the objective function. Defining such new constraints may not be overly complicated if done in two- or three-dimensional latent spaces, where the unexplored areas of the process can be easily visualized. However, doing so in higher-dimensional latent spaces may prove a challenge.
3. There is no standardized approach to tackle the problem of trying to maximize or minimize one or more outputs/quality attributes.
4. Furthermore, the optimization problem as formulated in the literature¹³ does not consider as hard constraints the bracketing of the DS¹⁴ or the subspace most likely to contain the DS, also referred to as ‘experiment space’¹⁶.

5. Lastly, none of the soft constraints in the objective function account for the effect of the leverage on the uncertainty in the DS estimation, i.e. any two given sets of scores located at the same Euclidean distance from the NS in the latent space will be equally penalized by the first term in the objective function.

These limitations will be addressed in the following sections.

3. Proposed formulation for the optimization in the latent space

While estimating the DS and solving an optimization problem are usually presented as different approaches to achieve a similar goal, there are some similarities between them that are worth considering. More importantly, identifying the facets of the envelope of the subspace of the KS most likely to contain the DS allows their use as hard constraints in the optimization problem.

Additionally, there is a very close relationship between the definition of the NS, as a necessary step to estimate the DS, and the objective function in the optimization problem. Regarding this, the first addend in the objective function in Eq. 2, $(\hat{\mathbf{y}}_{\text{NEW}} - \mathbf{y}_{\text{DES}})^T \cdot \mathbf{\Gamma} \cdot \mathbf{D}_{\text{SY}}^{-2} \cdot (\hat{\mathbf{y}}_{\text{NEW}} - \mathbf{y}_{\text{DES}})$, is a weighted sum of quadratic distances between the expected values of the outputs of the optimization solution and the corresponding desired values, with the l -th element in the diagonal matrix $\mathbf{\Gamma}$ being the weighting factor. Therefore, each of these quadratic distances is related to the distance between the optimization solution $\boldsymbol{\tau}$ and the NS corresponding to the desired value for the l -th output variable; they are not related to the distance to the subspace (if it exists) corresponding to all combinations of inputs theoretically leading to the desired values for all of the outputs. That is to say, the way the first addend in the objective function in Eq. 2 is formulated does not make use of the concept of the NS presented by Jaeckle & MacGregor¹¹ (also referred to as ‘combined pseudo-NS’ by García-Muñoz et al.¹²), but

(implicitly) to the L different NS as referred to by García-Muñoz et al.¹². Nonetheless, an equivalence exists between assessing the differences between the expected and desired values for each output, and evaluating the distance between $\boldsymbol{\tau}$ and each of the corresponding NSs. This relationship will be considered in the next sections for the reformulation of the optimization problem.

Note also that, if the NSs for the desired values for all outputs intersect, at least a solution meeting all the desired values for the L output variables, \mathbf{y}_{DES} , exists. If this intersection results in a subspace of dimension higher than zero (i.e. a line, plane or hyper-plane), then such intersection is coincident with, or very close to, the NS as defined by Jaeckle & MacGregor¹¹, or combined pseudo-NS by García-Muñoz et al.¹², depending on the number of latent variables used to fit the LVRM. The solution $\boldsymbol{\tau}$ of the optimization problem as formulated in Eq. 2 does not necessarily lead to \mathbf{y}_{DES} even in this scenario. As long as $g_1 > 0$ in Eq. 2, $\boldsymbol{\tau}$ will be ‘displaced’ from the intersection of all NSs in a direction that minimizes $T_{\boldsymbol{\tau}}^2$, even if some sets of scores in this subspace meet all of the hard constraints imposed on the solution of the optimization problem. This is because the greater g_1 is with respect to g_0 , the more $\boldsymbol{\tau}$ is ‘pushed’ towards the centre of projection and farther away from the desired values for the outputs.

3.1. Optimization of a linear combination of variables

3.1.1. Theoretical framework

The l -th NS can be defined as the subspace constituted by all combinations of inputs that, according to the PLS-regression model, guarantee the desired value $y_{\text{DES},l}$ for the l -th output variable, but not necessarily the rest, such that a $[L \times 1]$ vector $\hat{\mathbf{y}}_{\text{NS}_l}$, corresponding to a set of scores $\boldsymbol{\tau}_{\text{NS}_l}$ on the l -th NS, meets that¹⁶:

$$\mathbf{o}_l^T \cdot \hat{\mathbf{y}}_{NS_l} = \mathbf{o}_l^T \cdot (\mathbf{D}_{S_Y} \cdot \mathbf{Q} \cdot \boldsymbol{\tau}_{NS_l} + \mathbf{m}_Y) = y_{DES,l} \quad (3)$$

where \mathbf{o}_l is the $[L \times 1]$ vector whose l -th element is 1, and the rest are zeros.

Reorganizing terms, Eq. 3 provides the equation of a hyper-plane of the form:

$$\begin{aligned} v_{l,0} + \sum_{a=1}^A v_{l,a} \cdot \tau_{NS_l,a} &= 0 ; \mathbf{v}_l = [v_{l,1}, v_{l,2}, \dots, v_{l,A}]^T \\ v_{l,0} &= \mathbf{o}_l^T \cdot \mathbf{m}_Y - y_{DES,l} \\ \mathbf{v}_l &= \mathbf{Q}^T \cdot \mathbf{D}_{S_Y} \cdot \mathbf{o}_l \end{aligned} \quad (4)$$

And when applied to all L output variables:

$$\begin{aligned} \mathbf{v}_0 + \mathbf{V} \cdot \boldsymbol{\tau}_{NS} &= \mathbf{0} \\ \mathbf{v}_0 = \begin{bmatrix} v_{1,0} \\ v_{2,0} \\ \vdots \\ v_{L,0} \end{bmatrix} &= \mathbf{m}_Y - \mathbf{y}_{DES} ; \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_L^T \end{bmatrix} &= \mathbf{D}_{S_Y} \cdot \mathbf{Q} \end{aligned} \quad (5)$$

\mathbf{y}_{DES} being the $[L \times 1]$ column vector with $y_{DES,l}$ as its l -th element, and $\boldsymbol{\tau}_{NS}$ a set of scores in the intersection of the L NS (if it exists).

Note that, from Eq. 2 and Eq. 5, one can conclude that $\hat{\mathbf{y}}_{NEW} - \mathbf{y}_{DES} = \mathbf{v}_0 + \mathbf{V} \cdot \boldsymbol{\tau}$. More

generally, consider R quality attributes of interest, such that the r -th one is $d_r = \mathbf{a}_r^T \cdot \mathbf{y}$.

Then, for a set of scores $\boldsymbol{\tau}_{NS}$ in the intersection of the R NS associated to the desired values for all R quality attributes¹⁶:

$$\begin{aligned} \mathbf{v}_0 + \mathbf{V} \cdot \boldsymbol{\tau}_{NS} &= \mathbf{0} \\ \mathbf{v}_0 = \begin{bmatrix} v_{1,0} \\ v_{2,0} \\ \vdots \\ v_{R,0} \end{bmatrix} &= \mathbf{A}_R \cdot \mathbf{m}_Y - \mathbf{d}_{DES} ; \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_R^T \end{bmatrix} &= \mathbf{A}_R \cdot \mathbf{D}_{S_Y} \cdot \mathbf{Q} \end{aligned} \quad (6)$$

where \mathbf{A}_r is a $[R \times L]$ matrix with \mathbf{a}_r^T as its r -th row, and \mathbf{d}_{DES} a $[R \times 1]$ column vector whose r -th element is $d_{DES,r}$. Therefore, a more general formulation of Eq. 2 to account

for linear combinations of variables (in this case, quality attributes defined as linear combinations of outputs) can be seen in Eq. 7.

$$\begin{aligned}
& \min_{\boldsymbol{\tau}} [\text{OF}_1(g_{0a}, g_1, \boldsymbol{\tau})] \\
\text{OF}_1(g_{0a}, g_1, \boldsymbol{\tau}) &= g_{0a} \cdot (\mathbf{v}_0 + \mathbf{V} \cdot \boldsymbol{\tau})^T \cdot \boldsymbol{\Gamma}_R \cdot \mathbf{D}_{s_d}^{-2} \cdot (\mathbf{v}_0 + \mathbf{V} \cdot \boldsymbol{\tau}) + g_1 \cdot \sum_{a=1}^A \frac{\tau_a^2}{s_a^2} \\
\text{s. t.} & \\
\mathbf{v}_0 &= \mathbf{A}_r \cdot \mathbf{m}_Y - \mathbf{d}_{\text{DES}} \\
\mathbf{V} &= \mathbf{A}_r \cdot \mathbf{D}_{s_Y} \cdot \mathbf{Q} \\
\hat{\mathbf{y}}_{\text{NEW}} &= \mathbf{D}_{s_Y} \cdot \mathbf{Q} \cdot \boldsymbol{\tau} + \mathbf{m}_Y \\
\hat{\mathbf{x}}_{\text{NEW}} &= \mathbf{D}_{s_X} \cdot \mathbf{P} \cdot \boldsymbol{\tau} + \mathbf{m}_X \\
T_{\boldsymbol{\tau}}^2 &= \boldsymbol{\tau}^T \cdot \boldsymbol{\Lambda}^{-1} \cdot \boldsymbol{\tau} \leq T_{\text{lim}}^2 \\
\mathbf{A}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} &\leq \mathbf{d}_{\boldsymbol{\tau}} \\
\mathbf{F}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} &= \mathbf{f}_{\boldsymbol{\tau}}
\end{aligned} \tag{7}$$

where $\boldsymbol{\Gamma}_R$ is a $[R \times R]$ diagonal matrix where the r -th element in the diagonal represents the weight given to achieving the desired value for the r -th quality attribute (such that the sum of the absolute values of the diagonal elements is one); $\mathbf{D}_{s_d}^{-2}$ is the $[R \times R]$ diagonal matrix with $s_{d_r}^{-2}$ as its r -th element; and $s_{d_r}^2$ is calculated as in Eq. 8:

$$\begin{aligned}
s_{d_r}^2 &= \mathbf{a}_r^T \cdot \mathbf{S}_y \cdot \mathbf{a}_r \\
\mathbf{S}_y &= \begin{pmatrix} s_{y_1}^2 & \widehat{\text{cov}}(y_1, y_2) & \dots & \widehat{\text{cov}}(y_1, y_L) \\ \widehat{\text{cov}}(y_2, y_1) & s_{y_2}^2 & & \widehat{\text{cov}}(y_2, y_L) \\ \vdots & & \ddots & \vdots \\ \widehat{\text{cov}}(y_L, y_1) & \widehat{\text{cov}}(y_L, y_2) & \dots & s_{y_L}^2 \end{pmatrix} \\
s_{y_l}^2 &= \frac{\sum_{n=1}^N (y_{n,l} - \bar{y}_l)^2}{N - df} \\
\widehat{\text{cov}}(y_l, y_{l'}) &= \frac{\sum_{n=1}^N (y_{n,l} - \bar{y}_l) \cdot (y_{n,l'} - \bar{y}_{l'})}{N - df}
\end{aligned} \tag{8}$$

The relevance of this formulation lies in the fact that, in most cases, there is no direct equivalence between: i) optimizing (e.g. maximizing or minimizing, or achieving a desired value) d_r itself and; ii) optimizing the different y_l and defining $\boldsymbol{\Gamma}_R$ as a diagonal matrix with its diagonal somewhat related to \mathbf{a}_r (see Appendix).

Furthermore, as illustrated in one of the case studies presented by Palací-López et al.¹⁶, it is not always possible/practical to build a PLS-regression model that properly predicts

the quality attribute of interest, but it adequately predicts other output variables such quality attribute can be expressed as a linear combination of^{fii}.

To better illustrate the importance of properly formulating the optimization problem when linear combinations of variables are involved, consider the case study in the following section (Section 3.1.2).

3.1.2. Case Study 1

For this case study, data from a petrochemical process was used. The data set consists of a matrix with 2075 hourly measurements made on 135 input variables corresponding to processing conditions and 10 output variables related to the percentage of throughput of different products with respect to the total amount. Furthermore, all output variables are strongly correlated among each other and must meet a hard equality constraint that involves all of them, such that $\sum_{l=1}^{10} y_l = 1$ and $0 \leq y_l \leq 1 \forall l = 1, 2, \dots, 10$.

After a cross validation exercise to select the optimum number of latent variables, a PLS-regression model with $A = 14$ LVs was fitted capable of explaining $\sim 75.5\%$ of the variability of the inputs and $\sim 77.7\%$ of the variability of the outputs. In this case, the quality attribute of interest represents the expected revenue generated by the process per mass unit of products produced, and can be expressed as a linear combination of the ten outputs, $d = \mathbf{a}^T \cdot \mathbf{y}$ (where the prices in \mathbf{a} may vary on a daily or weekly basis), and its maximum observable value was $d_{\max_{\text{obs}}} = 171.58$.

In this example, 4 of the 135 inputs are required to meet specific values in the solution. Given these constraints, the maximum observed value for the quality attribute of interest

ⁱⁱⁱ Although not addressed in this paper, the extension of the formulation of the objective function in Eq. 7 and Eq. 8 to include input variables (and linear combinations of them) is straightforward.

was $d_{\max_{\text{obs}}}^* = 159.34$. Then, the optimization problem is solved following two distinct approaches:

- i. Using the formulation in Eq. 2, with $g_0 = g_1 = 1$, $y_{\text{DES},l} = 1 \forall l = 1, 2, \dots, 10$, and $\mathbf{\Gamma}$ having in its diagonal the a_l coefficients of the linear combination of output variables.
- ii. Using the formulation in Eq. 7, with $g_{0a} = g_1 = 1$, $d_{\text{DES}} = 175$ (i.e. above the maximum observed value in the whole historical dataset).

In both scenarios hard constraints are imposed so that the solution does not fall outside of the Hotelling T^2 95% confidence hyperellipsoid, the values for the input variables remain within historical limits, and the 4 of them which must take specific values satisfy these restrictions. Figure 1 shows the results from both approaches, with the observations coloured according to the value of the quality attribute of interest, expressed as $d = \mathbf{a}^T \cdot \mathbf{y}$.

[Figure 1]

The first thing to note from Figure 1 is that, as expected, the solutions obtained from both approaches (Eq. 2 and Eq. 7) are not the same. Furthermore, for the first approach, $\hat{d}_{\text{NEW}} = 160.577$ (black square in Figure 1), with the solution being slightly away from the cluster of high values for the quality attribute that can be seen in the upper right part of the ellipsoid, while for the second approach, $\hat{d}_{\text{NEW}} = 164.832$ (black circle in Figure 1), and the corresponding set of scores is located much closer to such cluster. This highlights the relevance of properly formulating the optimization problem when linear combinations of variables are involved in the objective function.

3.2. Optimization for exploration in the latent space

3.2.1. Theoretical framework

As mentioned in Section 2, the second limitation of the optimization formulation in Eq. 2 is the need to actively modify the soft and/or hard constraints (or the weight given to the former in the objective function). This is important in order to guarantee that different iterations will provide solutions different to previous ones, or to already available observations in the calibration dataset used to fit the PLS-regression model. Furthermore, when the optimization approach is used to perform new experiments (e.g., in unexplored areas of the KS), an additional issue must be considered: the first term in the objective function favours/penalizes equally acceptable solutions at the same distance (in the latent space) from the NS. However, the limits of the NS confidence region are non-linear (and, in particular and as an approximation, of the form presented by Palací-López et al.¹⁶) due to the prediction error varying with the leverage associated to each set of inputs/scores. Then it makes sense for exploratory purposes to penalize less solutions equally close to the NS, but with higher leverage, and vice versa.

When both of these concerns are accounted for, the following optimization problem can be formulated:

$$\begin{aligned}
& \min_{\boldsymbol{\tau}} [\text{OF}_2(g_{0a}, g_{0b}, g_1, g_2, \boldsymbol{\tau})] \\
\text{OF}_2(g_{0a}, g_{0b}, g_1, g_2, \boldsymbol{\tau}) &= \text{OF}_1(g_{0a}, g_1, \boldsymbol{\tau}) + g_{0b} \cdot \mathbf{d}_{\text{NS}}^T(\boldsymbol{\tau}, \text{NS}) \cdot \boldsymbol{\Gamma}_R \cdot \mathbf{d}_{\text{NS}}(\boldsymbol{\tau}, \text{NS}) + \\
& \quad + g_2 \cdot d^{-1}(\boldsymbol{\tau}, \mathbf{T}) \\
\text{s. t.} \\
\mathbf{v}_0 &= \mathbf{A}_r \cdot \mathbf{m}_Y - \mathbf{d}_{\text{DES}} \\
\mathbf{V} &= \mathbf{A}_r \cdot \mathbf{D}_{S_Y} \cdot \mathbf{Q} \\
\hat{\mathbf{y}}_{\text{NEW}} &= \mathbf{D}_{S_Y} \cdot \mathbf{Q} \cdot \boldsymbol{\tau} + \mathbf{m}_Y \\
\hat{\mathbf{x}}_{\text{NEW}} &= \mathbf{D}_{S_X} \cdot \mathbf{P} \cdot \boldsymbol{\tau} + \mathbf{m}_X \\
|d_{\text{NS}_r}(\boldsymbol{\tau}, \text{NS}_r)| &= \left| \frac{v_{r,0} + v_r^T \cdot \boldsymbol{\tau}}{s_{\hat{d}_{\text{NEW},r}}} \right| \leq t_{N-A, \frac{\alpha}{2}} \quad \forall r \in \{1, 2, \dots, R\} \\
T_{\boldsymbol{\tau}}^2 &= \boldsymbol{\tau}^T \cdot \boldsymbol{\Lambda}^{-1} \cdot \boldsymbol{\tau} \leq T_{\text{lim}}^2 \\
\mathbf{A}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} &\leq \mathbf{d}_{\boldsymbol{\tau}} \\
\mathbf{F}_{\boldsymbol{\tau}} \cdot \boldsymbol{\tau} &= \mathbf{f}_{\boldsymbol{\tau}}
\end{aligned} \tag{9}$$

where $\mathbf{\Gamma}_R$ holds the same meaning as in Eq. 7; g_{0b} and g_2 are the weights given to each of the new added terms in the objective function; $\mathbf{d}_{NS}(\boldsymbol{\tau}, \mathbf{NS})$ is a $[R \times 1]$ vector with $d_{NS_r}(\boldsymbol{\tau}, NS_r)$ as its r -th element, which is a measure of the distance from the projection $\boldsymbol{\tau}$ of the solution to the r -th NS weighted by the inverse of the estimated standard deviation of the r -th linear combination of outputs for $\hat{\mathbf{x}}_{NEW}$ (instead of the r -th diagonal element in $\mathbf{D}_{s_d}^{-2}$), $s_{\hat{d}_{NEW,r}}$, estimated as¹⁶:

$$s_{\hat{d}_{NEW,r}} = \sqrt{\mathbf{a}_r^T \cdot \mathbf{S}_f \cdot \mathbf{a}_r \cdot \left(1 + h_{\text{obs}} + \frac{1}{N}\right)}$$

$$\mathbf{S}_f = \begin{pmatrix} s_{f_1}^2 & \widehat{\text{cov}}(f_1, f_2) & \dots & \widehat{\text{cov}}(f_1, f_L) \\ \widehat{\text{cov}}(f_2, f_1) & s_{f_2}^2 & \dots & \widehat{\text{cov}}(f_2, f_L) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\text{cov}}(f_L, f_1) & \widehat{\text{cov}}(f_L, f_2) & \dots & s_{f_L}^2 \end{pmatrix} \quad (10)$$

$$s_{f_l}^2 = \frac{\sum_{n=1}^N (y_{n,l} - \hat{y}_{n,l})^2}{N - df}$$

$$\widehat{\text{cov}}(f_l, f_{l'}) = \frac{\sum_{n=1}^N (y_{n,l} - \hat{y}_{n,l}) \cdot (y_{n,l'} - \hat{y}_{n,l'})}{N - df}$$

and $d^{-1}(\boldsymbol{\tau}, \mathbf{T})$ is the inverse of the squared statistical distance from the projection $\boldsymbol{\tau}$ of the solution to the closest already available sample in the calibration dataset \mathbf{X} with projection \mathbf{T} in the latent subspace, calculated as:

$$d^{-1}(\boldsymbol{\tau}, \mathbf{T}) = \frac{1}{\min[(\boldsymbol{\tau} - \boldsymbol{\tau}_n)^T \cdot \boldsymbol{\Lambda}^{-1} \cdot (\boldsymbol{\tau} - \boldsymbol{\tau}_n)] + \text{const}} \quad \forall n \in \{1, 2, \dots, n\} \quad (11)$$

const being a small positive real number introduced in order to avoid $d^{-1}(\boldsymbol{\tau}, \mathbf{T})$ going to infinity, which may cause computational issues.

The hard constraints on $d_{NS_r}(\boldsymbol{\tau}, NS_r)$ guarantee that the solution will also fall within the confidence region (for a given confidence level) of the NS associated to the desired value for the r -th quality attribute, as defined by Palací-López et al. ¹⁶.

It must also be noted that the terms of the objective function weighted by g_{0a} and g_{0b} both penalize solutions farther away from the NS, but do so in different ways. At the same time, $d^{-1}(\boldsymbol{\tau}, \mathbf{T})$ guarantees that any solution obtained from the optimization will not be coincident or too close to an already available calibration observation.

One of the main disadvantages of this optimization approach is that the solution achieved may change substantially depending not only on the constraints imposed, but also on the weight given to each of the terms in the objective function $OF(g_{0a}, g_{0b}, g_1, g_2, \boldsymbol{\tau})$, even when there is only one quality attribute.

To illustrate this, consider the example in the following section (Section 3.2.2).

3.2.2. Case Study 2: general assessment of the new objective function

For this case study, consider the following polynomial model defining a hypothetical input-output causal relationship^{14,16}:

$$y = -21 + 4.3 \cdot x_1 + 0.022 \cdot x_2 - 0.0064 \cdot x_3 + 1.1 \cdot x_4 - 0.12 \cdot x_5 \quad (12)$$

$$\text{s. t. } x_3 = x_1^2 ; x_4 = x_2^2 ; x_5 = x_1 \cdot x_2$$

where y represents the ‘quality attribute of interest’ expressed as a function of the input variables x_1, x_2, x_3, x_4 and x_5 .

Six observations are generated following the model in Eq. 12, and according to a two-level full factorial design on x_1 and x_2 , with a replicate centre point. Random Gaussian noise with 5% of the variability of each variable in the dataset is then added to x_3, x_4, x_5 and y . Then, a PLS-regression model is fitted with 2 LVs.

Given the (arbitrarily defined) desired value for the output to be $y_{DES} = 204.86$, Figure 2 presents the values that the objective function in Eq. 9 takes when weight equal to 1 is given to one of the terms in it, and 0 to the rest, at each point in the latent space. This highlights the different contribution to each of the terms in the objective function, and

allows assessing the effect of different weights being given to each one (g_{0a} , g_{0b} , g_1 and g_2) prior to the optimization.

[Figure 2]

In Figures 2.a and 2.c the contributions of each term in the OF as formulated in Eq. 9 is assessed. It can be seen how these terms' contributions to the OF have similar order of magnitude if close to the centre of projection. Since during the optimization the lowest value for the OF is desired, the term corresponding to $OF(1,0,0,0)$ penalizes solutions farther away from the NS (for $y_{DES} = 204.86$), while the term associated to $OF(0,0,1,0)$ favours solutions with lower Hotelling T_{τ}^2 .

On the other hand, the contribution of the new proposed terms can be evaluated. Figure 2.b, with $OF(0,1,0,0)$, corresponds to $d_{NS_r}(\tau, NS_r)$. This term favours solutions close to the NS (in a similar manner as $OF(1,0,0,0)$), but penalizes solutions farther away from the NS much more heavily. Furthermore, solutions closer to points in the NS with a higher leverage (and therefore higher uncertainty in the prediction) are less penalized when compared to those close to points in the NS with a lower leverage, which implies that this term will allow a wider area for the exploration of different solutions where the predictions are less reliable. Figure 2.d assesses the impact of the new proposed term on the OF, through $OF(0,0,0,1)$, which penalizes solutions close to the projection onto the latent subspace of already available observations in the calibration dataset.

Figure 2.e corresponds to $OF(1,0,1,1)$, the sum of the terms plotted in Figure 2.a, Figure 2.c and Figure 2.d. Since the contributions of $OF(1,0,0,0)$ and $OF(0,0,1,0)$ are of a similar order of magnitude close to the centre of projection, there are solutions farther away from the NS that are equally optimal (with this formulation of the OF) to solutions closer to it if g_{0a} and g_1 are given the same values. Finally, Figure 2.f corresponds to

OF(0,1,1,1), that is, the sum of the terms plotted in Figure 2.b, Figure 2.c and Figure 2.d. In this case any solution will fall, even without hard restrictions, closer to the NS than if the weights given to each term were those corresponding to Figure 2.e, even if g_{0b} and g_1 are given the same values. This serves to illustrate the importance of giving the appropriate weight to each term in the objective function to guarantee that more desirable solutions are achieved

3.2.3. Case Study 2 (cont): solving the optimization problem with different weighting terms.

To better illustrate the usefulness of the new terms, consider the same dataset simulated in the previous section 3.2.2. Then, the optimization formulation in Eq. 9 is used ten consecutive times to obtain ten different sets of inputs, given $y_{DES} = 204.86$, hard constraints on the inputs so that their values in the solution are not below the minimum observed ones nor above the maximum observed ones in the calibration dataset, and a hard constraint on the Hotelling T_{τ}^2 so that it does not fall outside of the Hotelling T_{τ}^2 confidence hyperellipsoid. Four variations of the optimization formulation in Eq. 9 are used as enumerated in Table 1:

[Table 1]

Figure 3 shows the results from these four scenarios.

[Figure 3]

Solving the optimization problem when $g_2 = 0$ more than once (even accounting for previous results) will always return almost indistinguishable solutions (apparently, a single circle in both plots). However, a different one will be obtained each time the problem is solved when $g_2 = 1$ (ten different squares in both plots).

When comparing Figure 3.a with Figure 3.b, on the other hand, it is possible to assess the difference in using $g_{0a} = 1$ and $g_{0b} = 0$ (Figure 3.a) vs $g_{0a} = 0$ and $g_{0b} = 1$ (Figure 3.b). While in the first case the weight of the term associated to the Hotelling T_{τ}^2 is of a similar magnitude to the term associated to achieving y_{DES} , in the second case the Hotelling T_{τ}^2 term does not influence the solution obtained as much. As a consequence, one of the solutions in Figure 3.a for $g_2 = 1$ is far enough from the NS to be located outside of its confidence region (and, as a consequence, outside of the experimental region¹⁶, shown in grey in Figure 3), while this is not the case in Figure 3.b. Also, in Figure 3.a the solution(s) of the optimization problem for $g_2 = 0$ is closer to the centre of projection than to the NS, while the opposite is true in Figure 3.b.

Finally, Figure 3 also illustrates how a portion of the NS (the segment within the grey area) meets all of the hard constraints imposed on the solution. This also highlights the importance of assessing:

- i) if a more convenient formulation of the optimization problem would require dismissing the terms associated to soft constraints on the quality attributes in the objective function (i.e. $g_{0a} = g_{0b} = 1$) and imposing, instead, hard constraints on them, such that the solution will be exactly on the NS and with smallest leverage (and therefore a narrower prediction confidence interval for the prediction of the quality attribute/s);
- ii) if the prediction error is not of concern at this stage, thus considering all the combinations of inputs/scores on the NS as ‘equally good’ solutions (which would, in fact, require defining this subspace instead of solving the optimization problem).

3.3. Tackling the minimization/maximization problem

3.3.1. Theoretical framework

As illustrated in Figure 2, the first term in the objective function of the optimization problem as formulated in Eq. 9 penalizes solutions far away from the hyperplanes defined by the NS corresponding to a vector of ‘desired values’ for the considered outputs/quality attributes. This may present some issues when such ‘desired value’ is not a specific one, but instead the value for the corresponding quality attribute is to be minimized or maximized.

One possible approach may consist on defining such ‘desired values’ as the minimum/maximum feasible ones if known, or as arbitrarily low/high values otherwise. To illustrate why such approach may not be the most appropriate, consider a scenario for which the information in Table 2 is available concerning two hypothetical output variables y_1 and y_2 , for their average, minimum and maximum values and their standard deviation in the calibration dataset used to fit the PLS-regression model, as well as their minimum and maximum feasible values.

[Table 2]

Consider now three different scenarios:

1. Both y_1 and y_2 are to be maximized. If the vector of desired values for the outputs is defined as $\mathbf{y}_{\text{DES}}^T = [1000 \quad 120]$, the difference between $\mathbf{y}_{\text{DES},1}$ and the maximum observed value for y_1 is ~ 556 standard deviations, while the difference between $\mathbf{y}_{\text{DES},2}$ and the maximum observed value for y_2 is less than 3 standard deviations. Therefore, defining $\mathbf{\Gamma}_R = \mathbf{I}_2$, \mathbf{I}_2 being the $[2 \times 2]$ identity matrix, makes it look like the same importance is being given to achieve the ‘desired’ (i.e. maximum) value for both outputs. However, accomplishing this goal for the first output is being given ~ 189 times as much weight as achieving it

for the second one. This will only *not* be a problem when the combinations of inputs (e.g. process conditions) that maximize both y_1 and y_2 are the same.

2. Both y_1 and y_2 are to be minimized. In this case, if the vector of desired values for the outputs is defined as $\mathbf{y}_{DES}^T = [30 \ 0]$, the difference between $y_{DES,1}$ and the minimum observed value for y_1 is ~ 8.8 standard deviations, while the difference between $y_{DES,2}$ and the minimum observed value for y_2 is ~ 26.5 standard deviations. In contrast to the previous scenario, when $\mathbf{\Gamma}_R = \mathbf{I}_2$ the goal of minimizing y_2 is being given roughly 3 times the weight given to minimizing y_1 . Again, this will only *not* be a problem when the combinations of inputs that minimize both y_1 and y_2 are the same.
3. y_1 is to be minimized/maximized, and y_2 is to take a specific desired value, or vice versa. A similar issue to that in the two previous cases occurs here, to a higher or lower extent depending on which variable is to be minimized/maximized, and the desired value for the other one.

It can then be concluded that this way of defining \mathbf{y}_{DES} (or \mathbf{d}_{DES}) may be inadequate in most cases, especially if a feasible minimum or maximum is not known or defining one does not make sense from a practical point of view (in which case an arbitrarily low or high value would be set). To understand how common these scenarios can be, consider that y_1 and y_2 may represent a temperature, pressure, mass or volume of production, or yield, just to mention some examples. It is easy to see how these variables present values that are usually orders of magnitude apart from each other on average, or whose variability significantly differ from one another.

The proposed approach takes into account that any solution of the optimization problem as formulated in Eq. 9 may be not valid (optimality aside). In particular, no solution

outside of the Hotelling T^2 confidence hyperellipsoid, for a given confidence level, will be accepted, independently of any other restrictions imposed on the inputs or outputs or the feasible minimum or maximum values for any variable. Therefore, not accounting for additional constraints, the lowest and highest predicted values for a quality attribute according to the PLS-regression model will be those for two sets of scores located on opposite extremes of the Hotelling T^2 confidence hyperellipsoid. To identify them, consider that the vector \mathbf{v}_r orthogonal to the NS associated to any given value of the r -th quality attribute, and defined as the transpose of the r -th row of \mathbf{V} in Eq. 6, provides the direction of maximum variability for such attribute^{iv}. Consider also that the equation of the Hotelling T^2 confidence hyperellipsoid can be written as:

$$\boldsymbol{\tau}^T \cdot \frac{\boldsymbol{\Lambda}^{-1}}{T_{\text{lim}}^2} \cdot \boldsymbol{\tau} = 1 \quad (13)$$

And, $\boldsymbol{\Lambda}^{-1}$ being a diagonal matrix, the vector normal to such hyperellipsoid on a point $\boldsymbol{\tau}_r^+$ on its surface, $\mathbf{n}_{\boldsymbol{\tau}_r^+}$, is the same as the vector orthogonal to the hyperplane that is tangent to such hyperellipsoid at that point²³. Therefore:

$$\mathbf{n}_{\boldsymbol{\tau}_r^+} = \frac{2}{T_{\text{lim}}^2} \cdot \boldsymbol{\Lambda}^{-1} \cdot \boldsymbol{\tau}_r^+ \quad (14)$$

If $\boldsymbol{\tau}_r^+$ is the set of scores associated to the combination of inputs theoretically (i.e. according to the PLS-regression model) guaranteeing the maximum achievable value for the r -th quality attribute without reaching outside of the Hotelling T^2 hyperellipsoid for a given confidence level, then $\boldsymbol{\tau}_r^+$ is a point that is located simultaneously on such hyperellipsoid, and on the NS associated to the maximum achievable value for the r -th quality attribute, which will be the hyperplane tangent to the hyperellipsoid at $\boldsymbol{\tau}_r^+$.

^{iv} As discussed in ¹⁶, since \mathbf{v}_r is a vector orthogonal to the r -th NS, it is also orthogonal to all directions of null variability for the r -th quality attribute, and therefore provides its direction of maximum variability

Therefore, $\mathbf{n}_{\tau_r^+}$ and \mathbf{v}_r will be two vectors that are parallel to each other and, as a consequence:

$$\begin{aligned} \frac{\mathbf{n}_{\tau_r^+}}{\sqrt{\mathbf{n}_{\tau_r^+}^T \cdot \mathbf{n}_{\tau_r^+}}} &= \frac{\mathbf{v}_r}{\sqrt{\mathbf{v}_r^T \cdot \mathbf{v}_r}} \\ \tau_r^{+T} \cdot \frac{\Lambda^{-1}}{T_{\text{lim}}^2} \cdot \tau_r^+ &= 1 \end{aligned} \quad (15)$$

On the other hand, if τ_r^- is the set of scores associated to the combination of inputs theoretically guaranteeing the minimum achievable value for the r -th quality attribute without reaching outside of the Hotelling T^2 hyperellipsoid for a given confidence level, then:

$$\left. \begin{aligned} \frac{\mathbf{n}_{\tau_r^-}}{\sqrt{\mathbf{n}_{\tau_r^-}^T \cdot \mathbf{n}_{\tau_r^-}}} &= \frac{-\mathbf{v}_r}{\sqrt{\mathbf{v}_r^T \cdot \mathbf{v}_r}} \\ \tau_r^{-T} \cdot \frac{\Lambda^{-1}}{T_{\text{lim}}^2} \cdot \tau_r^- &= 1 \end{aligned} \right\} \rightarrow \mathbf{n}_{\tau_r^-} = -\mathbf{n}_{\tau_r^+} \rightarrow \tau_r^- = -\tau_r^+ \quad (16)$$

Eq. 15 and Eq. 16 can then be solved to obtain τ_r^+ and τ_r^- , and from them the maximum and minimum achievable values for the r -th quality attribute while inside the Hotelling T^2 confidence hyperellipsoid, respectively. The PLS-regression model is then used to predict the values for the outputs or quality attributes of interest for the sets of scores τ_r^- and τ_r^+ . These values can be used as the r -th element of \mathbf{d}_{DES} in the optimization problem formulated in Eq. 9. While using them does not necessarily guarantee that balanced weightings are implicitly given to the goals of maximizing/minimizing each of the quality attributes, the procedure to define them does not depend on any “feasibility constraints” or on the decision of the person resorting to the optimization problem as formulated in Eq. 9 to give them arbitrarily large/small values. Incidentally, if a single quality attribute is considered for the optimization, and if τ_r^+ or τ_r^- meet all other hard

constraints imposed on the solution of the optimization algorithm, these scores can be considered the solution of the algorithm itself when no soft constraint is imposed on its Hotelling T_{τ}^2 (i.e. $g_1 = 0$).

3.3.2. Case Study 3

To better illustrate the impact of defining the ‘desired values’ one way or the other (i.e. setting them to the minimum/maximum feasible ones vs using the proposed approach), data were simulated following the procedure explained by Arteaga & Ferrer^{24,25} to generate multivariate normal data. Fifty observations with six variables (four considered as inputs (x_i) and two as outputs (y_i)) following the correlation structure shown in Table 3¹⁶ were simulated.

[Table 3]

Then, the data corresponding to variables y_1 and y_2 were transformed to have mean and standard deviation equal to the hypothetical dataset in Table 2. For graphical representation purposes a PLS-regression model fitted with two LVs ($R^2(\mathbf{X},2LV)=0.642$; $Q^2(\mathbf{Y},2LV)=0.858$) is chosen. Scenarios 1 and 2 presented below Table 2 are solved first using as $y_{DES,1}$ and $y_{DES,2}$ the maximum (for scenario 1) and minimum (for scenario 2) feasible values shown in Table 2, and then with the approach presented in this paper. Figure 4 presents the different results obtained in both cases.

[Figure 4]

Figure 4 shows clearly what is expected to happen when using the minimum (or maximum) feasible values for the corresponding output variables, i.e. the solution being conditioned by the differences in variability and ‘desired values’ for each variable with respect to the minimum or maximum ones in the calibration dataset. On the other hand, using the proposed approach to define \mathbf{y}_{DES} to tackle the minimization/maximization

problem leads to intermediate solutions that do not minimize/maximize any of the two outputs, which makes sense given that the weight given to achieving such values is the same (i.e. $\mathbf{\Gamma}_R = \mathbf{I}_2$) and both goals compete with each other. Therefore, this approach guarantees that the importance given to getting the ‘desired values’ for the minimization/maximization problem remains as indicated by $\mathbf{\Gamma}_R$.

4. Conclusions

This paper focused on three important limitations of the PLS-regression model-based optimization approach in the latent space. Namely, the lack of a formulation that allows a standardized approach to:

1. Optimizing any quality attribute that can be expressed as a linear combination of outputs.
2. Obtaining new solutions in subsequent optimizations and/or solutions that span previously unexplored areas of the knowledge space without relying on changes in the hard or soft constraints imposed on these solutions (on the inputs, outputs, or any combination of them).
3. Tackling the issue of trying to minimize/maximize one or more quality attributes using the quadratic formulation without unintentionally giving too much/too little weight to doing so for one or more quality attributes with respect to others.

As illustrate by the results presented, the proposed formulation of the optimization problem allows efficiently solving all of these limitations, which are present in some of the most similar formulations in the literature¹³.

Furthermore, the inclusion of both soft and hard constraints related to the confidence region of the NS is also proposed, mainly for exploratory purposes, in order to acknowledge the increase in the uncertainty in the definition of the NS with the leverage

along its extension. The optimization problem, formulated this way, may help in obtaining a wider range of sets of inputs that guarantee the desired values for the quality attributes of interest (or close to them) in a systematic way. While not addressed in this paper, and left for future work, such approach may help to improve the estimation of the DS by gathering more samples within the subspace of the KS where the DS is expected to be located.

As with past formulations in the literature, the one proposed in this paper resorts to quadratic programming and will occasionally suffer from issues such as non-convergence and the presence of local minima. Future work will attempt to alleviate these limitations.

Acknowledgements

This study resulted from the collaboration between the Computer-Aided Process Engineering (CAPE-Lab) group at the *Università degli Studi di Padova* and the Multivariate Statistical Engineering Group (MSEG) at the *Universitat Politècnica de València* (UPV). Financial support was granted by the UPV as a part of the Erasmus+ Programme: Key Action 1 – Mobility for learners and staff – Higher Education Student and Staff Mobility, and was also supported by the Spanish Ministry of Economy and Competitiveness under the project DPI2017-82896-C2-1-R MINECO/AEI/FEDER, UE.

Appendix. Optimization of a linear combination of outputs

Consider any set of output variables $\hat{\mathbf{y}}_{NS,r}$ for which the desired value for the r -th quality attribute of interest, d_r , is met, that is $\mathbf{a}_r^T \cdot \hat{\mathbf{y}}_{NS,r} = d_{DES,r}$. The first term in the objective function in the optimization problem as formulated by Tomba et al.¹³ is expressed as a function of $\hat{\mathbf{y}}_{NEW}$ and \mathbf{y}_{DES} , and not $\mathbf{a}_r^T \cdot \hat{\mathbf{y}}_{NEW}$ and $d_{DES,r}$ (see Eq. 2). To evaluate the

adequacy of this formulation when the r -th quality attribute is taken into account, instead of one or more outputs, consider the following objective function:

$$\min_{\boldsymbol{\tau}} \left[g_0 \cdot (\mathbf{a}_r^T \cdot \hat{\mathbf{y}}_{\text{NEW}} - d_{\text{DES},r})^T \cdot \gamma_r \cdot s_{d_r}^{-2} \cdot (\mathbf{a}_r^T \cdot \hat{\mathbf{y}}_{\text{NEW}} - d_{\text{DES},r}) + g_1 \cdot \sum_{a=1}^A \frac{\tau_a^2}{s_a^2} \right] \quad (\text{A.1})$$

Where γ_r is a scalar that represents the weight given to achieving the desired value for the r -th quality attribute of interest, such that the sum of the absolute values for the weights given to achieving the desired values for each quality attribute is one (if only one quality attribute is considered, as in this case, then $\gamma_r = 1$); and $s_{d_r}^{-2}$ is the inverse of the estimated variance for the r -th quality attribute from the observations in the calibration dataset.

For the objective functions in Eq. 2 and Eq. A.1 to be equivalent, the first term in each of them must be so. This first term in Eq. A.1 can be reformulated as:

$$(\hat{\mathbf{y}}_{\text{NEW}} - \hat{\mathbf{y}}_{\text{NS},r})^T \cdot \gamma_r \cdot \mathbf{a}_r \cdot s_{d_r}^{-2} \cdot \mathbf{a}_r^T \cdot (\hat{\mathbf{y}}_{\text{NEW}} - \hat{\mathbf{y}}_{\text{NS},r}) \quad (\text{A.2})$$

where $\hat{\mathbf{y}}_{\text{NS},r}$ is any set of outputs that meets that $\mathbf{a}_r^T \cdot \hat{\mathbf{y}}_{\text{NS},r} = d_{\text{DES},r}$. When $\hat{\mathbf{y}}_{\text{NS},r}$ in Eq. A.2 coincides with \mathbf{y}_{DES} in Eq. 2, then:

$$\boldsymbol{\Gamma} \cdot \mathbf{D}_{s\mathbf{y}}^{-2} = \gamma_r \cdot \mathbf{a}_r \cdot s_{d_r}^{-2} \cdot \mathbf{a}_r^T \quad (\text{A.3})$$

It must be noted, however, that the estimated variance for the quality attribute from the observations in the calibration dataset, $s_{d_r}^2$, is calculated as:

$$s_{d_r}^2 = \mathbf{a}_r^T \cdot \mathbf{S}_y \cdot \mathbf{a}_r$$

$$\mathbf{S}_y = \begin{pmatrix} s_{y_1}^2 & \widehat{\text{cov}}(y_1, y_2) & \dots & \widehat{\text{cov}}(y_1, y_L) \\ \widehat{\text{cov}}(y_2, y_1) & s_{y_2}^2 & \dots & \widehat{\text{cov}}(y_2, y_L) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\text{cov}}(y_L, y_1) & \widehat{\text{cov}}(y_L, y_2) & \dots & s_{y_L}^2 \end{pmatrix} \quad (\text{A.4})$$

$$s_{y_l}^2 = \frac{\sum_{n=1}^N (y_{n,l} - \bar{y}_l)^2}{N - df}$$

$$\widehat{\text{cov}}(y_l, y_{l'}) = \frac{\sum_{n=1}^N (y_{n,l} - \bar{y}_l) \cdot (y_{n,l'} - \bar{y}_{l'})}{N - df}$$

\bar{y}_l and $y_{n,l}$ being, respectively, the average and n -th measured value for the l -th output of the model calibration dataset, and df the degrees of freedom consumed by the model. Therefore:

$$\mathbf{\Gamma} = \gamma_r \cdot \mathbf{a}_r \cdot (\mathbf{a}_r^T \cdot \mathbf{S}_y \cdot \mathbf{a}_r)^{-1} \cdot \mathbf{a}_r^T \cdot \mathbf{D}_{s_y}^2 \quad (\text{A.5})$$

Since $\mathbf{\Gamma}$ in the objective function in Eq. 2 is defined as a diagonal matrix of weightings, the objective function in Eq. 2 and the one formulated in Eq. A.1 will only be equivalent, with $\mathbf{\Gamma}$ remaining a diagonal matrix, if the outputs are uncorrelated^v.

References

- (1) FDA. *Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach*; 2004.
- (2) Juran JM. *Juran on Quality by Design: The New Steps for Planning Quality Into Goods and Services*; Schuster, Simon, Eds.; Free Press, 1992.
- (3) ICH. Pharmaceutical Development Q8. *ICH Harmon. Tripart. Guidel.* 2009;8(August):1–28.
- (4) Liu JJ, MacGregor JF. Modeling and Optimization of Product Appearance:

^v The extension of this demonstration when all R quality attributes of interest are considered is straightforward

- Application to Injection-Molded Plastic Panels. *Ind. Eng. Chem. Res.* 2005;**44**(13):4687–4696.
- (5) Bonvin D, Georgakis C, Pantelides CC, Barolo M, Grover MA, Rodrigues D, Schneider R, Dochain D. Linking Models and Experiments. *Ind. Eng. Chem. Res.* 2016;**55**(25):6891–6903.
- (6) Box GEP, Hunter JS, Hunter WG. *Statistics for Experimenters: Design, Discovery, and Innovation*, Second.; John Wiley & Sons, INC., 2005.
- (7) Montgomery DC. *Applied Statistics and Probability for Engineers Third Edition*; 2003; Vol. 37.
- (8) MacGregor JF. Empirical Models for Analyzing “Big” Data-What’s the Difference. In *Spring AIChE Conference*; Orlando, Florida, USA, 2018.
- (9) Liu Z, Bruwer M-J, F. MacGregor J, Rathore S, E. Reed D, J. Champagne M. Modeling and Optimization of a Tablet Manufacturing Line. *J. Pharm. Innov.* 2011;**6**:170–180.
- (10) MacGregor JF, Bruwer MJ, Miletic I, Cardin M, Liu Z. Latent Variable Models and Big Data in the Process Industries. *IFAC-PapersOnLine* 2015;**28**(8):520–524.
- (11) Jaeckle CM, MacGregor JF. Industrial Applications of Product Design through the Inversion of Latent Variable Models. *Chemom. Intell. Lab. Syst.* 2000;**50**(2):199–210.
- (12) García-Muñoz S, Kourti T, MacGregor JF, Apruzzese F, Champagne M. Optimization of Batch Operating Policies. Part I. Handling Multiple Solutions. *Ind. Eng. Chem. Res.* 2006;**45**(23):7856–7866.
- (13) Tomba E, Barolo M, García-Muñoz S. General Framework for Latent Variable

- Model Inversion for the Design and Manufacturing of New Products. *Ind. Eng. Chem. Res.* 2012;**51**(39):12886–12900.
- (14) Facco P, Dal Pasto F, Meneghetti N, Bezzo F, Barolo M. Bracketing the Design Space within the Knowledge Space in Pharmaceutical Product Development. *Ind. Eng. Chem. Res.* 2015;**54**(18):5128–5138.
- (15) Bano G, Facco P, Bezzo F, Barolo M. Probabilistic Design Space Determination in Pharmaceutical Product Development: A Bayesian/Latent Variable Approach. *AIChE J.* 2018;**64**(7):2438–2449.
- (16) Palací-López D, Facco P, Barolo M, Ferrer A. New Tools for the Design and Manufacturing of New Products Based on Latent Variable Model Inversion. *Chemom. Intell. Lab. Syst.* 2019;**194**:103848.
- (17) MacGregor JF, Bruwer M-J. A Framework for the Development of Design and Control Spaces. *J. Pharm. Innov.* 2008;**3**(1):15–22.
- (18) Jaeckle C, Macgregor J. Product Design through Multivariate Statistical Analysis of Process Data. *Comput. Chem. Eng.* 1996;**20**(96):S1047–S1052.
- (19) Lakshminarayanan S, Fujii H, Grosman B, Dassau E, Lewin DR. New Product Design via Analysis of Historical Databases. *Comput. Chem. Eng.* 2000;**24**:671–676.
- (20) García-Muñoz S, MacGregor JF, Neogi D, Latshaw BE, Mehta S. Optimization of Batch Operating Policies. Part II. Incorporating Process Constraints and Industrial Applications. *Ind. Eng. Chem. Res.* 2008;**47**(12):4202–4208.
- (21) Wold S, Sjöström M, Eriksson L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* 2001;**58**:109–130.
- (22) Ferrer A. Multivariate Statistical Process Control Based on Principal Component

- Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Qual. Eng.* 2007;**19**(4):311–325.
- (23) Feltens J. Vector Method to Compute the Cartesian (X , Y , Z) to Geodetic (φ , λ , h) Transformation on a Triaxial Ellipsoid. *J. Geod.* 2009;**83**:129–137.
- (24) Arteaga F, Ferrer A. Building Covariance Matrices with the Desired Structure. *Chemom. Intell. Lab. Syst.* 2013;**127**:80–88.
- (25) Arteaga F, Ferrer A. How to Simulate Normal Data Sets with the Desired Correlation Structure. *Chemom. Intell. Lab. Syst.* 2010;**101**(1):38–42.

Table 1. Case Study 2 (cont): values given to the weighting terms
in the objective function in four optimization problem scenarios

	g_{0a}	g_{0b}	g_1	g_2
i)	1	0	1	0
ii)	1	0	1	1
iii)	0	1	1	0
iv)	0	1	1	1

Table 2. *Characterization of a hypothetical \mathbf{Y} dataset with two output variables*

Variable	Dataset				Feasible	
	Mean	Std. dev	Minimum	Maximum	Minimum	Maximum
y_1	50	1.7	45	55	30	1000
y_2	95	3.4	90	110	0	120

Table 3. *Case Study 3: Correlation matrix used to generate the dataset*

	x_1	x_2	x_3	x_4	y_1
x_2	0.45				
x_3	0.54	0.50			
x_4	0.30	0.55	0.70		
y_1	0.50	-0.10	-0.15	0.20	
y_2	-0.30	0.34	0.50	-0.15	-0.70