



HAL
open science

Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume

Patrick J C Tardivel, Rémi Servien, Didier Concordet

► To cite this version:

Patrick J C Tardivel, Rémi Servien, Didier Concordet. Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume. 2018. hal-01904873v1

HAL Id: hal-01904873

<https://hal.science/hal-01904873v1>

Preprint submitted on 25 Oct 2018 (v1), last revised 6 Jul 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume

Patrick J.C. Tardivel^{1*}, Rémi Servien² and Didier Concordet²

¹ Institute of Mathematics, Wrocław University, Wrocław, Poland.

² INTHERES, Université de Toulouse, INRA, ENVT, Toulouse, France.

Abstract

We study the control of the FamilyWise Error Rate (FWER) in the linear Gaussian model when the $n \times p$ design matrix is of rank p . Single step multiple testing procedures controlling the FWER are derived from hyperrectangular confidence regions. In this study, we aim to construct procedures derived from hyperrectangular confidence regions having a minimal volume. We show that minimizing the volume seems a fair criterion to improve the power of the multiple testing procedure. Numerical experiments demonstrate the performance of our approach when compared with the state-of-the-art single step and sequential procedures. We also provide an application to the detection of metabolites in metabolomics.

Keywords: Confidence region, Multiple testing procedure, Metabolomics

1 Introduction

Let us consider the linear Gaussian model

$$Y = X\beta^* + \varepsilon, \tag{1}$$

where $X = (X_1 | \dots | X_p)$ is an $n \times p$ design matrix of rank p with $p < n$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 Id_n)$, and $\beta^* \in \mathbb{R}^p$ is an unknown parameter. We aim to test the hypotheses $\mathcal{H}_i : \beta_i^* = 0$, with $1 \leq i \leq p$. Several type I errors can be controlled in such multiple testing procedures. In this study, we focus on the Familywise Error Rate (FWER), defined as the probability of wrongly rejecting at least one hypothesis \mathcal{H}_i . Let $\hat{\beta}^{\text{mle}} := (X'X)^{-1}X'Y$ be the maximum likelihood estimator of the model (1). The usual multiple testing procedures are based on the maximum likelihood estimator and reject $\mathcal{H}_i : \beta_i^* = 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s$, where $s \geq 0$ is the same threshold for all the hypotheses $\mathcal{H}_1, \dots, \mathcal{H}_p$. Let ζ be a random vector having the same distribution as

*corresponding author: tardivel@math.uni.wroc.pl

$(\hat{\beta}_1^{\text{mle}}/\text{se}(\hat{\beta}_1^{\text{mle}}), \dots, \hat{\beta}_p^{\text{mle}}/\text{se}(\hat{\beta}_p^{\text{mle}}))$ when $\beta^* = \mathbf{0}$. We observe that ζ is a Gaussian vector or a multivariate student, depending on whether the standard errors $\text{se}(\hat{\beta}_1^{\text{mle}}), \dots, \text{se}(\hat{\beta}_p^{\text{mle}})$ are known or estimated (thus, σ is known or estimated). There are several ways to choose such a threshold s , assuring control of the FWER at a significance level $\alpha \in (0, 1)$. For example, s can be chosen according to correlation-free inequalities such as the Bonferroni inequality [Dunn, 1959] or the Gaussian correlation inequality¹ [Royen, 2014, Šidák, 1967] (available when ζ is Gaussian)

$$\mathbb{P}\left(\bigcup_{1 \leq i \leq p} \{|\zeta_i| \geq s\}\right) \leq \sum_{i=1}^p \mathbb{P}(|\zeta_i| \geq s) \text{ and } \mathbb{P}(|\zeta_1| \geq s, \dots, |\zeta_p| \geq s) \geq \prod_{i=1}^p \mathbb{P}(|\zeta_i| \geq s).$$

These inequalities are extremely convenient and provide a threshold for controlling the FWER at the significance level α . The first and second inequalities, respectively, provide the thresholds $s_{\text{bonf}} := q_{1-\alpha/2p}$ and $s_{\text{sidak}} := q_{(1+\sqrt{1-\alpha})/2}$, where q_η denotes the η quantile of a ζ_1 . By taking into account the correlation, a smaller threshold s_{max} (thus, a better power) is given by setting s_{max} as the $1 - \alpha$ quantile of $\max\{|\zeta_1|, \dots, |\zeta_p|\}$.

Confidence regions and testing procedures are closely related (see *e.g.* [Lehmann and Romano, 2005] page 72). Historically, the famous Bonferroni and Dunn-Šidák corrections for multiple testing procedures [Dunn, 1959, Šidák, 1967] originated from the construction of hyperrectangular confidence regions (also called simultaneous confidence intervals). Actually, taking $s \in \{s_{\text{bonf}}, s_{\text{sidak}}, s_{\text{max}}\}$ gives the hyperrectangular confidence region $[\hat{\beta}_1^{\text{mle}} \pm s \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \dots \times [\hat{\beta}_p^{\text{mle}} \pm s \times \text{se}(\hat{\beta}_p^{\text{mle}})]$, which contains β^* with a probability larger than $1 - \alpha$. Conversely, given a hyperrectangular confidence region $[\hat{\beta}_1^{\text{mle}} \pm s_1 \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \dots \times [\hat{\beta}_p^{\text{mle}} \pm s_p \times \text{se}(\hat{\beta}_p^{\text{mle}})]$ containing β^* with a probability larger than $1 - \alpha$ provides a procedure by rejecting $\mathcal{H}_i : \beta_i^* = 0$ when $0 \notin [\hat{\beta}_i^{\text{mle}} \pm s_i \times \text{se}(\hat{\beta}_i^{\text{mle}})]$ (or equivalently, by rejecting \mathcal{H}_i when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_i$). This last procedure controls the FWER at significance level α . We argue that none of the thresholds $s_{\text{bonf}}, s_{\text{sidak}}$ and s_{max} provides a hyperrectangular confidence region with the smallest volume.

It is natural to attempt minimizing the volume for a confidence region. We aim to illustrate that deriving a multiple testing procedure from a hyperrectangular confidence region having a minimal volume is an intuitive way to improve power. In addition, we aim to present a new operational procedure through a numerical method for volume minimization. This article is organized as follows. Section 2 contains some basic properties about the optimal hyperrectangular confidence region. We exhibit some cases in which it is convenient to perform the computation of the optimal hyperrectangular confidence region.

Section 3 presents a method to numerically minimize the volume of the hyperrectangular confidence region.

Section 4 is devoted to simulation experiments: we compare our multiple testing procedure with the state-of-the-art single step and sequential procedures.

¹The inequality $\mathbb{P}(|\zeta_1| \geq s_1, \dots, |\zeta_p| \geq s_p) \geq \prod_{i=1}^p \mathbb{P}(|\zeta_i| \geq s_i)$ already proved in Šidák [1967] is a particular case of the Gaussian correlation inequality $\mathbb{P}(|\zeta_1| \geq s_1, \dots, |\zeta_p| \geq s_p) \geq \mathbb{P}(|\zeta_i| \geq s_1, \dots, |\zeta_i| \geq s_k) \mathbb{P}(|\zeta_i| \geq s_{k+1}, \dots, |\zeta_i| \geq s_p)$. Recently, the Gaussian correlation inequality was proved by Royen [2014].

Section 5 details the analysis of metabolomic data, which motivated this study.

2 Minimization of volume

We use the following notations:

- The sets \mathcal{A}_0 and \mathcal{A}_1 are, respectively, $\mathcal{A}_0 := \{i \in \{1, \dots, p\} \mid \beta_i^* = 0\}$ and $\mathcal{A}_1 := \{i \in \{1, \dots, p\} \mid \beta_i^* \neq 0\}$.
- The matrix $M_p(a, b)$ is a $p \times p$ matrix for which the diagonal elements are a and the non-diagonal elements are b .
- Given a random vector $V := (V_1, \dots, V_p)$, $\text{var}(V)$ denotes the covariance matrix of V and $\text{var}(V_i)$ denotes the marginal variance of V_i .
- The matrix Σ is a $p \times p$ semi-definite positive matrix and C is a $p \times p$ ‘correlation’ matrix, namely, C is a semi-definite positive matrix, such that $C_{11} = \dots = C_{pp} = 1$.
- Note that ζ is a random vector having the same distribution as $(\hat{\beta}_1^{\text{mle}}/\text{se}(\hat{\beta}_1^{\text{mle}}), \dots, \hat{\beta}_p^{\text{mle}}/\text{se}(\hat{\beta}_p^{\text{mle}}))$ when $\beta^* = \mathbf{0}$. Consequently, depending on whether or not $\text{se}(\hat{\beta}_1^{\text{mle}}), \dots, \text{se}(\hat{\beta}_p^{\text{mle}})$ are known or estimated, $\zeta \sim \mathcal{N}(\mathbf{0}, C)$ or $\zeta \sim t_{n-p}(\mathbf{0}, C)$ where C is a ‘correlation’ matrix.

We aim to construct a multiple testing procedure derived from a hyperrectangular confidence region for β^* having the following expression: $[\hat{\beta}_1^{\text{mle}} \pm s_1 \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \dots \times [\hat{\beta}_p^{\text{mle}} \pm s_p \times \text{se}(\hat{\beta}_p^{\text{mle}})]$. To guarantee a significance level of $1 - \alpha$ (with $\alpha \in (0, 1)$) the thresholds s_1, \dots, s_p must satisfy the following inequality:

$$\mathbb{P}(\beta^* \in [\hat{\beta}_1^{\text{mle}} \pm s_1 \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \dots \times [\hat{\beta}_p^{\text{mle}} \pm s_p \times \text{se}(\hat{\beta}_p^{\text{mle}})]) = 1 - \alpha \Leftrightarrow \mathbb{P}(|\zeta_1| \leq s_1, \dots, |\zeta_p| \leq s_p) = 1 - \alpha.$$

Among these hyperrectangular confidence containing β^* with a probability larger than $1 - \alpha$, we aim to pick one for which the volume $2^p s_1 \times \dots \times s_p$ is minimal, which leads to the following optimisation problem:

$$\text{minimize } \prod_{i=1}^p s_i \text{ subject to } \mathbb{P}(|\zeta_1| \leq s_1, \dots, |\zeta_p| \leq s_p) = 1 - \alpha. \quad (2)$$

Thus far, the components of ζ are ‘standardized’ and the correlation matrix is invertible (since $X'X$ is invertible). We will show that the construction of the hyperrectangular confidence region (thus, the derived multiple testing procedure) does not depend on this standardization. The problem (2) is still meaningful when ζ has an arbitrary covariance matrix Σ (a relevant example for Σ is $\sigma^2(X'X)^{-1}$ the covariance matrix of $\hat{\beta}^{\text{mle}}$), and just for the proposition 1, we assume that $\zeta \sim \mathcal{N}(\mathbf{0}, \Sigma)$ or $\zeta \sim t_{n-p}(\mathbf{0}, \Sigma)$, where Σ is a positive semi-definite $p \times p$ matrix (thus, Σ not invertible is allowed in this proposition). The proposition 1 shows that there exist optimal thresholds s_1^*, \dots, s_p^* solution of (2) and there is no loss of generality to restrict our attention to the particular case in which $\Sigma_{11} = \dots = \Sigma_{pp} = 1$.

Proposition 1 Let Σ be a $p \times p$ positive semi-definite matrix such that $\Sigma_{11} \neq 0, \dots, \Sigma_{pp} \neq 0$ and let ζ be a Gaussian vector $\mathcal{N}(\mathbf{0}, \Sigma)$ or a multivariate student $t_{n-p}(\mathbf{0}, \Sigma)$. Then, the following properties hold

1. **Existence of optimal thresholds:** There exists at least one element $s^* \in [0, +\infty)^p$, which is the solution of the problem (2).
2. **Standardization:** If s^* is a solution of the problem (2), then $(s_1^*/\sqrt{\Sigma_{11}}, \dots, s_p^*/\sqrt{\Sigma_{pp}})$ is a solution of the following problem:

$$\text{minimize } \prod_{i=1}^p s_i \text{ subject to } \mathbb{P} \left(\frac{|\zeta_1|}{\sqrt{\Sigma_{11}}} \leq s_1, \dots, \frac{|\zeta_p|}{\sqrt{\Sigma_{pp}}} \leq s_p \right) = 1 - \alpha. \quad (3)$$

We do not need the uniqueness of the minimizer of (2), but only pick a particular optimal threshold s^* . Criteria other than volume, such as the maximal length of the simultaneous confidence intervals ($\max\{s_1, \dots, s_p\}$) [Benjamini et al., 2013], could be investigated. However, contrarily to the volume, this last criterion does not provide a minimizer satisfying the standardization property. Indeed, if s^∞ and \bar{s}^∞ , respectively, are the solutions of problems (2) and (3) when $\prod_{i=1}^p s_i$ is substituted by $\max\{s_1, \dots, s_p\}$ then $s_1^\infty = \dots = s_p^\infty$ and $\bar{s}_1^\infty = \dots = \bar{s}_p^\infty$, and consequently, $\bar{s}^\infty \neq (s_1^\infty/\sqrt{\Sigma_{11}}, \dots, s_p^\infty/\sqrt{\Sigma_{pp}})$. Thus, the procedure rejecting $\beta_i^* = 0$ when $|\hat{\beta}_i^{\text{mle}}| > s^\infty$ is not equivalent to the procedure rejecting $\beta_i^* = 0$ when $|\hat{\beta}_i^{\text{mle}}|/\sqrt{\Gamma_{ii}} > \bar{s}^\infty$. As we do not intend to present a procedure that depends on the standardization of $\hat{\beta}^{\text{mle}}$, from a practical perspective, we do not use criteria for which the standardization property does not hold.

Henceforth, we assume that the components of ζ are ‘standardized’, and thus, $\zeta \sim \mathcal{N}(0, C)$ or $\zeta \sim t_{n-p}(0, C)$, where C is a ‘correlation’ matrix. In the Gaussian framework, given a minimizer s^* , the following proposition holds:

Proposition 2 Let ζ be a Gaussian vector $\mathcal{N}(0, C)$ with C an invertible correlation matrix, let $s^* = (s_1^*, \dots, s_p^*)$ be a solution of the optimisation problem (2). Let T^{s^*} denote the truncated Gaussian vector on $S^* = [-s_1^*, s_1^*] \times \dots \times [-s_p^*, s_p^*]$ having the following density:

$$f_{T^{s^*}}(u) = \frac{1}{(1 - \alpha)\sqrt{(2\pi)^p \det(C)}} \exp\left(-\frac{1}{2}uC^{-1}u\right) \mathbb{1}_{u \in S^*}$$

then all the diagonal coefficients of $C^{-1}\text{var}(T^{s^*})$ are equal.

Note that if the covariance matrix of T^{s^*} (here denoted by $\text{var}(T^{s^*})$) was equal to C , all the diagonal coefficients of $C^{-1}\text{var}(T^{s^*})$ would be equal, indicating that s^* is a solution of (2). Because the diagonal terms of $\text{var}(T^{s^*})$ are always smaller than the diagonal terms of C , $\text{var}(T^{s^*})$ cannot be equal to C . However, the condition given by proposition 2 can be intuitively interpreted. The optimal (with respect to the volume) hyperrectangular should be such that the covariance of the truncated Gaussian vector ζ restrained to $[-s_1^*, s_1^*] \times \dots \times [-s_p^*, s_p^*]$ is

as close as possible to the non-constraint covariance of the random vector ζ . The Gaussian framework has some simple yet interesting cases where the computation of the optimal thresholds s_1^*, \dots, s_p^* can be performed by hand. Note that in the special case $p = 2$ (i.e. where $\hat{\beta}^{\text{mle}}$ has only two components), basic algebra shows that, whatever the correlation between the two components, the optimal hyperrectangular is an hypercube. This property does not hold true when $p > 2$. Let us give the optimal thresholds in the following three examples. For convenience, we denote $M_p(a, b)$, a $p \times p$ matrix whose diagonal coefficients are equal to a and whose non-diagonal coefficients are equal to b .

1) In the independent case: Let us set $C = Id_p$. From the proposition 2, the vector s^* must satisfy

$$\text{var}(T_1^{s^*}) = \dots = \text{var}(T_p^{s^*}).$$

Let $c_{1-\alpha}$ be the $1 - \alpha$ quantile of the random variable $\max\{|\zeta_1|, \dots, |\zeta_p|\}$. The unique minimizer of the problem (1) is s^* , where $s_1^* = \dots = s_p^* = c_{1-\alpha}$.

2) In the equicorrelated case: Let us set $C = M_p(1, \rho)$, it follows that $C^{-1} = M_p(a, b)$ for some a and b .

Let $c_{1-\alpha}$ be the $1 - \alpha$ quantile of the random variable $\max\{|\zeta_1|, \dots, |\zeta_p|\}$ and let us set s^* such that $s_1^* = \dots = s_p^* = c_{1-\alpha}$. Because $\text{var}(T^{s^*}) = M(c, d)$ for some c and d , all the diagonal coefficients of $C^{-1}\text{var}(T^{s^*}) = M(a, b)M(c, d)$ are equal; thus, s^* is a minimizer of (1).

3) In the block diagonal equicorrelated case: Let us set $C = \text{diag}(M_k(1, \rho), M_{p-k}(1, \rho'))$. It follows that

C^{-1} is block diagonal with $C^{-1} = \text{diag}(M_k(a, b), M_{p-k}(a', b'))$ for some a, b, a', b' . If we set $s_1^* = \dots = s_k^* = c_1$ and $s_{k+1}^* = \dots = s_p^* = c_2$, one deduces that $\text{var}(T^{s^*})$ is block diagonal with $\text{var}(T^{s^*}) = \text{diag}(M(c, d), M(c', d'))$ for some c, d, c', d' . Consequently, whatever c_1 and c_2 , the k first diagonal coefficients of $C^{-1}\text{var}(T^{s^*})$ are equal and the $p - k$ last diagonal coefficients of $C^{-1}\text{var}(T^{s^*})$ are equal. We only need to tune c_1 and c_2 such that all diagonal coefficients of $C^{-1}\text{var}(T^{s^*})$ become equal.

As an illustration of the proposition 2, the thresholds prescribed in Dunnett's procedure are optimal.

Example: Dunnett's procedure compares the mean of treatment groups with the mean of the control group [Dunnett, 1955]. When each group (control and treatment) has the same number of observations, the proposition 2 shows that the thresholds given by Dunnett's procedure are the optimal ones.

Let us denote p as the number of treatment groups and n as the number of observations, under the assumptions of Dunnett (Gaussianity, homoscedasticity), the empirical means $\hat{M}_0, \hat{M}_1, \dots, \hat{M}_p$ of each group are independent and distributed according to $\mathcal{N}(\mu_0, \sigma^2/n), \mathcal{N}(\mu_1, \sigma^2/n), \dots, \mathcal{N}(\mu_p, \sigma^2/n)$ distributions (with σ known to simplify).

In Dunnett's procedure, whatever $i \in \{1, \dots, p\}$, the null hypothesis $\mathcal{H}_i^0 : \mu_i = \mu_0$ is rejected for the alternative $\mathcal{H}_i^1 : \mu_i \neq \mu_0$ as soon as $|\hat{M}_i - \hat{M}_0| > c_{1-\alpha} \sigma \sqrt{2/n}$, where $c_{1-\alpha}$ is the $1 - \alpha$ quantile of $\max\{|\zeta_1|, \dots, |\zeta_p|\}$.

In the later expression, $\zeta := (\zeta_1, \dots, \zeta_p)$ is distributed according to $\mathcal{N}(\mathbf{0}, \mathbf{C})$, where C is a $p \times p$ equicorrelated correlation matrix defined by $C_{ij} = 1$ if $i = j$ and $C_{ij} = 1/2$ otherwise. Because C is an equicorrelated correlation matrix, the threshold prescribed in Dunnett's procedure (the same for each hypothesis) is optimal with respect to the volume.

Except for these three particular cases, we do not have a closed form for the optimal thresholds. Therefore, we develop a numerical method to compute these optimal thresholds efficiently.

3 Numerical solver for the optimal thresholds

The optimal thresholds are provided by the solution of the following problem (equivalent to (2))

$$\min f(s) = \sum_{i=1}^p \ln(s_i) \text{ subject to } F(s) = \mathbb{P}(|\zeta_1| \leq s_1, \dots, |\zeta_p| \leq s_p) = 1 - \alpha.$$

Let $u \in (0, +\infty)^p$, the notation $1/u$ denotes $(1/u_1, \dots, 1/u_p)$. The Lagrange multiplier theorem assures that at s^* , the minimizer of (2), the vector $\nabla f(s^*)$ is collinear to $\nabla F(s^*)$, where ∇ denotes the gradient. Consequently, the following equivalences hold:

$$\frac{1}{s^*} \propto \nabla F(s^*) \Leftrightarrow s^* \propto \frac{1}{\nabla F(s^*)} \Leftrightarrow s^* \propto s^* + \frac{1}{\nabla F(s^*)} \text{ where } u \propto v \text{ means that } u \text{ is collinear to } v.$$

This collinearity motivates us to consider the following iterative sequence:

Let us set $s^{(0)} = (c_{1-\alpha}, \dots, c_{1-\alpha})$, where $c_{1-\alpha}$ is the $1 - \alpha$ quantile of $\max\{|\zeta_1|, \dots, |\zeta_p|\}$ and let us define the iterative sequence $(s^{(i)})_{i \in \mathbb{N}}$, where $s^{(i+1)}$ is given by

$$\begin{cases} u^{(i)} = \left(s^{(i)} + \frac{1}{\nabla F(s^{(i)})} \right), \\ s^{(i+1)} = \lambda_{1-\alpha} u^{(i)} \text{ where } \lambda_{1-\alpha} \text{ is the } 1 - \alpha \text{ quantile of } \max \left\{ \frac{|\zeta_1|}{u_1^{(i)}}, \dots, \frac{|\zeta_p|}{u_p^{(i)}} \right\}. \end{cases}$$

In the previous expression, because $\nabla F(s^{(i)}) \geq 0$, then $u^{(i)} \geq s^{(i)}$ (these two inequalities are given as per component). The parameter $\lambda_{1-\alpha}$ shrinks $u^{(i)}$ in order to recover an element $s^{(i+1)}$ so that $\mathbb{P}(|\zeta_1| \leq s_1^{(i+1)}, \dots, |\zeta_p| \leq s_p^{(i+1)}) = 1 - \alpha$. Thus far, this numerical method is available whatever ζ having a continuous distribution on \mathbb{R}^p and a covariance matrix Σ . However, the naive computation of the gradient $\nabla F(s^{(i)})$ through simulation is time expensive. Roughly, the components of $s^{(i)}$ are large; thus, $\mathbb{P}(s_j^{(i)} \leq |\zeta_j| \leq s_j^{(i)} + h)$ is very small, and consequently, a good estimation of $(F(s_1, \dots, s_{j-1}, s_j + h, s_{j+1}, \dots, s_p) - F(s_1, \dots, s_p)) / h$ through simulations is very time consuming. Fortunately, there is a trick to compute ∇F in the Gaussian and student frameworks. For example, let us explain how this trick provides the first component of ∇F . When ζ is Gaussian, the condi-

tional distribution $\mathcal{L}(\zeta_1 | \zeta_2 = x_2, \dots, \zeta_p = x_p)$ is a Gaussian distribution with density $f_{m(x), \sigma^2}$ (the mean $m(x)$ depends on $x := (x_2, \dots, x_p)$ while the variance σ^2 does not depend on x). Precisely, let $A = (C_{1j})_{2 \leq j \leq p}$ and let $B = (C_{ij})_{2 \leq i, j \leq p}$, then $m(x) = C_{12}B^{-1}x$ and $\sigma^2 = C_{11} - AB^{-1}A'$. Let $f_{\zeta_1}(\cdot | |\zeta_2| \leq s_2, \dots, |\zeta_p| \leq s_p)$ be the density of $\mathcal{L}(\zeta_1 | |\zeta_2| \leq s_2, \dots, |\zeta_p| \leq s_p)$ and let $f_{\zeta_2, \dots, \zeta_p}$ be the density of $(\zeta_2, \dots, \zeta_p)$, the first component of ∇F is given hereafter

$$\frac{\partial F}{\partial s_1}(s) = 2f_{\zeta_1}(s_1 | |\zeta_2| \leq s_2, \dots, |\zeta_p| \leq s_p) = 2 \int_{-s_2}^{s_2} \dots \int_{-s_p}^{s_p} f_{\zeta_2, \dots, \zeta_p}(x) f_{m(x), \sigma^2}(s_1).$$

Finally, the last expression is easy to compute by simulation. Notice that this trick requires C to be invertible. However, when C is no longer invertible, one can use the solver with the correlation matrix $\tilde{C} = (eId_p + C)/(1+e)$, where $e > 0$. The matrix \tilde{C} is invertible and when e is very small the optimal thresholds associated with \tilde{C} are almost equal to those associated with C . When $\zeta \sim t_{n-p}(\mathbf{0}, C)$ then ζ has the same distribution as $Z/\sqrt{V/(n-p)}$, where $Z \sim \mathcal{N}(\mathbf{0}, C)$ is independent of $V \sim \chi_d^2$ with $d = n - p$. In the student framework, the first component of ∇F is given hereafter

$$\frac{\partial F}{\partial s_1}(s) = 2f_{\zeta_1}(s_1 | |\zeta_2| \leq s_2, \dots, |\zeta_p| \leq s_p) = \int_0^{+\infty} 2f_{Z_1}\left(s_1\sqrt{\frac{v}{d}} \mid |Z_2| \leq s_2\sqrt{\frac{v}{d}}, \dots, |Z_p| \leq s_p\sqrt{\frac{v}{d}}\right) f_V(v)dv.$$

This integral is still easy to infer through simulation, by adding, with respect to the Gaussian framework, simulations of V . Now, in the Gaussian framework, we illustrate the performance of our solver using two examples.

Setting 1: We set $C := \text{diag}(M_{500}(1, 0.9), Id_{500})$. Because C is block diagonal equicorrelated, the optimal thresholds satisfy $s_1^*, \dots, s_{500}^* = c_1$ and $s_{501}^*, \dots, s_{1000}^* = c_2$ where c_1 and c_2 are handily computable. Thus, in this setting, it is easy to compare the theoretical optimal thresholds with the thresholds given by the solver of the problem (2).

Setting 2: We set $C = (C_{ij})_{1 \leq i, j \leq 1000}$ with $C_{ij} = \min\{i, j\} / \max\{i, j\}$. The matrix C is the correlation matrix of a Brownian motion discretized on the set $\{1, \dots, 1000\}$.

The left panel of the figure 1 shows that in setting 1 after $i = 5$ iterations, the threshold $s^{(i)}$ almost recovers the optimal thresholds. The right panel shows that in setting 2, $s_1^{(5)} \geq \dots \geq s_{999}^{(5)}$ (there is a singularity for $s_{1000}^{(5)}$ that is not a numerical problem).

As described in the figure 1, in both settings, the gain between the volume of the hypercube associated with the initial threshold $s^{(0)}$ (for which $s_1^{(0)} = \dots = s_{1000}^{(0)}$) and that of the hyperrectangular region associated with the threshold $s^{(5)}$ is very large.

Given optimal thresholds s_1^*, \dots, s_p^* , the solution of the problem (2), one derives a multiple testing procedure for the null hypotheses $\mathcal{H}_i : \beta_i^* = 0, i \in \{1, \dots, p\}$. The hypothesis \mathcal{H}_i is rejected when $0 \notin [\hat{\beta}_i^{\text{mle}} \pm s_i^* \times \text{se}(\hat{\beta}_i^{\text{mle}})]$

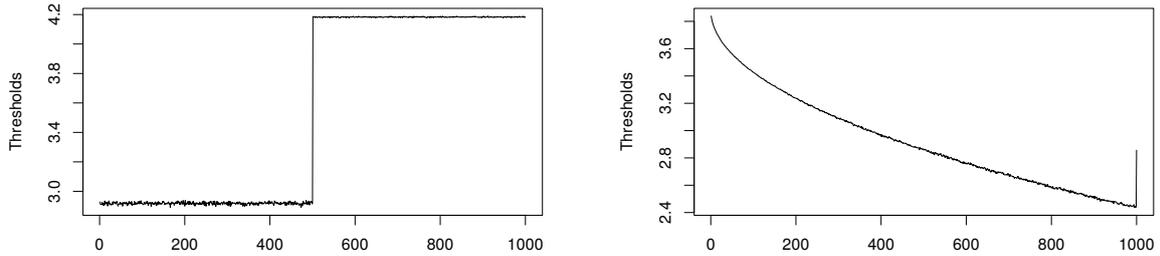


Figure 1: The figure on the left provides the optimal thresholds associated with setting 1 described above. A handy computation of the optimal thresholds gives $s_1^* = \dots = s_{500}^* = 2.94$ and $s_{501}^* = \dots = s_{1000}^* = 4.16$. The y-axis of the figure provides the thresholds $s_1^{(5)}, \dots, s_{1000}^{(5)}$ given by the iterative sequence $s^{(i)}$ after $i = 5$ iterations. Observe that our solver almost recovers the optimal thresholds. With respect to the initial and classical threshold $s^{(0)}$, the gain in volume is very large, as $\sum_{j=1}^{1000} \log(s_j^{(0)}) = 1360.650$ while $\sum_{j=1}^{1000} \log(s_j^{(5)}) = 1251.031$. The figure on the right provides the optimal thresholds associated with setting 2 described above. The y-axis of the figure provides the thresholds $s_1^{(5)}, \dots, s_{1000}^{(5)}$ given by the iterative sequence $s^{(i)}$ after $i = 5$ iterations. Again, the gain in volume is also large in setting 2, as $\sum_{j=1}^{1000} \log(s_j^{(0)}) = 1124.615$ while $\sum_{j=1}^{1000} \log(s_j^{(5)}) = 1064.401$.

or equivalently, when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_i^*$. Because the confidence region $[\hat{\beta}_1^{\text{mle}} \pm s_1^* \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \dots \times [\hat{\beta}_p^{\text{mle}} \pm s_p^* \times \text{se}(\hat{\beta}_p^{\text{mle}})]$ contains β^* with a probability larger than $1 - \alpha$, the previous procedure controls the Family-Wise Error Rate (FWER) at significance level α . Note that $\mathcal{A}_0 := \{i \in \{1, \dots, p\} \mid \beta_i^* = 0\}$, the FWER is the probability $\mathbb{P}\left(\bigcup_{i \in \mathcal{A}_0} |\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_i^*\right)$. The following inequality assures the control of the FWER at significance level α .

$$\mathbb{P}\left(\bigcup_{i \in \mathcal{A}_0} \frac{|\hat{\beta}_i^{\text{mle}}|}{\text{se}(\hat{\beta}_i^{\text{mle}})} > s_i^*\right) = \mathbb{P}\left(\bigcup_{i \in \mathcal{A}_0} \beta_i^* \notin [\hat{\beta}_i^{\text{mle}} \pm s_i^* \times \text{se}(\hat{\beta}_i^{\text{mle}})]\right) \leq \mathbb{P}\left(\bigcup_{i=1}^p \beta_i^* \notin [\hat{\beta}_i^{\text{mle}} \pm s_i^* \times \text{se}(\hat{\beta}_i^{\text{mle}})]\right) \leq \alpha.$$

Intuitively, because the volume of the hyperrectangular confidence region is minimal, one should expect to recover a multiple testing more powerful than the classical single step procedures. The numerical experiments given in the following section confirm this intuition.

4 Comparison of multiple testing procedures

The more famous single-step procedures controlling the FWER are Bonferroni's procedure [Dunn, 1961] (hereafter, procedure 1) and the procedure described in Lehmann and Romano [2005] page 352 (hereafter, procedure 2).

Procedure 1: Whatever $i \in \{1, \dots, p\}$, the null hypothesis $\mathcal{H}_i : \beta_i^* = 0$ is rejected in favour of the alternative $\beta_i^* \neq 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_{\text{bonf}}$. The threshold s_{bonf} is the $1 - \alpha/(2p)$ quantile of a ζ_1 .

Procedure 2: Whatever $i \in \{1, \dots, p\}$, the null hypothesis $\mathcal{H}_i : \beta_i^* = 0$ is rejected in favour of the alternative

$\beta_i^* \neq 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_{\max}$. The threshold s_{\max} is the $1 - \alpha$ quantile of $\max\{|\zeta_1|, \dots, |\zeta_p|\}$.

Sequential procedures have better power than single step procedures, especially when β^* has many large components. Hereafter, we describe the StepDown (SD) counterpart of procedures 1 and 2.

The generic stepdown procedure defined by Romano and Wolf [2005], Lehmann and Romano [2005] p. 352 and Dudoit and Van Der Laan [2007] p. 126 is a generalization of Holm's sequential procedure [Holm, 1979] (hereafter, procedure 1 SD). To describe the generic stepdown procedure, let us denote $T_i = \hat{\beta}_i^{\text{mle}}/\text{se}(\hat{\beta}_i^{\text{mle}})$. The statistical tests are sorted from the most significant to the least significant, namely, $|T_{r(1)}| \geq \dots \geq |T_{r(p)}|$. The rejection of the hypotheses $\mathcal{H}_{r(1)}, \dots, \mathcal{H}_{r(p)}$ is done sequentially, as explained hereafter.

Procedure 1 SD: The hypothesis $\mathcal{H}_{r(1)}$ is rejected if $|T_{r(1)}| > q_{1-\alpha/(2p)}$. The hypothesis $\mathcal{H}_{r(2)}$ is rejected if $|T_{r(1)}| > q_{1-\alpha/(2p)}$ and $|T_{r(2)}| > q_{1-\alpha/(2(p-1))}$ and so on. In the previous expressions, q_η is the η quantile of ζ_1 .

Procedure 2 SD: The hypothesis $\mathcal{H}_{r(1)}$ is rejected if $|T_{r(1)}| \geq t_{r(1)}$. The hypothesis $\mathcal{H}_{r(2)}$ is rejected if $|T_{r(1)}| > t_{r(1)}$ and $|T_{r(2)}| > t_{r(2)}$ and so on. In the previous expressions, the threshold $t_{r(s)}$ is the $1 - \alpha$ quantile of $\max\{|\zeta_{r(s)}|, \dots, |\zeta_{r(p)}|\}$.

We will compare procedures 1, 2, 1 SD and 2 SD with procedure 3, described hereafter, which is derived from the computation of the optimal thresholds.

Procedure 3: Whatever $i \in \{1, \dots, p\}$, the null hypothesis $\mathcal{H}_i : \beta_i^* = 0$ is rejected in favour of the alternative $\beta_i^* \neq 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_i^*$. The thresholds s_1^*, \dots, s_p^* are the optimal ones given in (2).

By construction of the thresholds $s_{\text{bonf}}, s_{\max}$ and s_1^*, \dots, s_p^* , the single step procedures 1, 2 and 3 control the FWER at a significance level $\alpha \in (0, 1)$. In addition, the procedures 1 SD and 2 SD also control the FWER at a significance level $\alpha \in (0, 1)$ (see *e.g.* Lehmann and Romano [2005] pages 351 to 353). A comparison of these five procedures based on the average power is carried out on the following setting:

The Gaussian vector ζ has a $\mathcal{N}(\mathbf{0}, C)$ distribution, where $C := \text{diag}(M_{900}(1, \rho), Id_{100})$. We set $\beta^* \in \mathbb{R}^{1000}$ with $\text{card}(\mathcal{A}_1) = k \in \{50, 500\}$ and for all $i \in \mathcal{A}_1, \beta_i^* = t$, where $t > 0$. For the different values of $\rho \in \{0, 0.5, 0.9, 1\}$, the optimal thresholds s_1^*, \dots, s_{1000}^* given by the proposition 1 are as follows.

- When $\rho = 0$ then $s_1^* = \dots = s_{900}^* = c_1 \approx 4.0553$ and $s_{901}^* = \dots = s_{1000}^* = c_2 \approx 4.0553$.
- When $\rho = 0.5$ then $s_1^* = \dots = s_{900}^* = c_1 \approx 3.7628$ and $s_{901}^* = \dots = s_{1000}^* = c_2 \approx 4.0961$.
- When $\rho = 0.9$ then $s_1^* = \dots = s_{900}^* = c_1 \approx 2.9284$ and $s_{901}^* = \dots = s_{1000}^* = c_2 \approx 4.3327$.
- When $\rho = 1$ then $s_1^* = \dots = s_{900}^* = c_1 \approx 1.9950$ and $s_{901}^* = \dots = s_{1000}^* = c_2 \approx 4.4061$.

Let $I_1 = \{1, \dots, 900\}$, let $I_2 = \{901, 1000\}$ and let $V \sim \mathcal{N}(t, 1)$. In this framework, the average power of procedure 3 is

$$\frac{\text{card}(\mathcal{A}_1 \cap I_1)}{\text{card}(\mathcal{A}_1)} \mathbb{P}_t(|V| > c_1) + \frac{\text{card}(\mathcal{A}_1 \cap I_2)}{\text{card}(\mathcal{A}_1)} \mathbb{P}_t(|V| > c_2).$$

We observe that the average power of procedure 3 depends on the location of the non-null components of β^* . Intuitively, when the non-null components of β^* are located on I_1 (*i.e.* $\mathcal{A}_1 \subset I_1$) then, because the thresholds s_1^*, \dots, s_{900}^* are small, procedure 3 should be powerful. On the other hand, when the non-null components of β^* are located on I_2 (*i.e.* $\mathcal{A}_1 \subset I_2$) then, because the thresholds $s_{901}^*, \dots, s_{1000}^*$ are large, procedure 3 should not be powerful. Thus, to perform a fair comparison, instead of computing the average power for a particular \mathcal{A}_1 , we will examine the expected value of the average power when ‘ \mathcal{A}_1 ’ is a random set uniformly distributed on the set of combination of k elements among 1000. Let U be a random set with a uniform distribution on the set $\{I \subset \{1, \dots, 1000\} \mid \text{card}(I) = k\}$, the expected value of the average power of procedure 3 is given hereafter

$$\mathbb{E}_U \left(\frac{\text{card}(U \cap I_1)}{k} \mathbb{P}_t(|V| > c_1) + \frac{\text{card}(U \cap I_2)}{k} \mathbb{P}_t(|V| > c_2) \right) = \frac{9}{10} \mathbb{P}_t(|V| > c_1) + \frac{1}{10} \mathbb{P}_t(|V| > c_2). \quad (4)$$

In the figure 2, we compare the average power of procedures 1, 2, 3, 1 SD and 2 SD.

A stepdown procedure is merely a sequential application of a single step procedure. Precisely, in the first stage of stepdown procedure 1 SD (resp. 2 SD), the rejections are the ones given by the single step procedure 1 (resp. 2). The second stage of stepdown procedure 1 SD (resp. 2 SD) is an application of single step procedure 1 (resp. 2) on the hypotheses not rejected in the first stage, and so on. Intuitively, when β^* has many very large components, a large number of hypotheses are rejected in the first stage, implying that the number of hypotheses tested in the second stage becomes small, allowing the stepdown procedure to become powerful.

The most favourable setting for the stepdown procedures is when $\text{card}(\mathcal{A}) = 500$ and t is large, as in this case, there exist a large number of large components for β^* . Note that our procedure is at least as competitive as the other ones in this situation, and depending on ρ , our procedure can be much more powerful than the state-of-the art procedures.

In summary, when $\hat{\beta}^{\text{mle}}$ has correlated components, our procedure should always give better results (with respect to the average power) than all existing ones. When some components of $\hat{\beta}^{\text{mle}}$ are highly correlated, if β^* is sparse (*i.e.* has a large number of null components) our method outperforms the other ones. In the non-sparse and slightly correlated setting, stepdown procedure 2 SD gives mildly better results than ours.

It could appear as appealing to construct a stepdown procedure based on the procedure 3. Unfortunately, as illustrated in the appendix the application of the generic stepdown method on the procedure 3 (as described in Lehmann and Romano [2005] page 353) does not control the FWER. The construction of a stepdown procedure based on the procedure 3 and controlling the FWER is, for the authors, an open question.

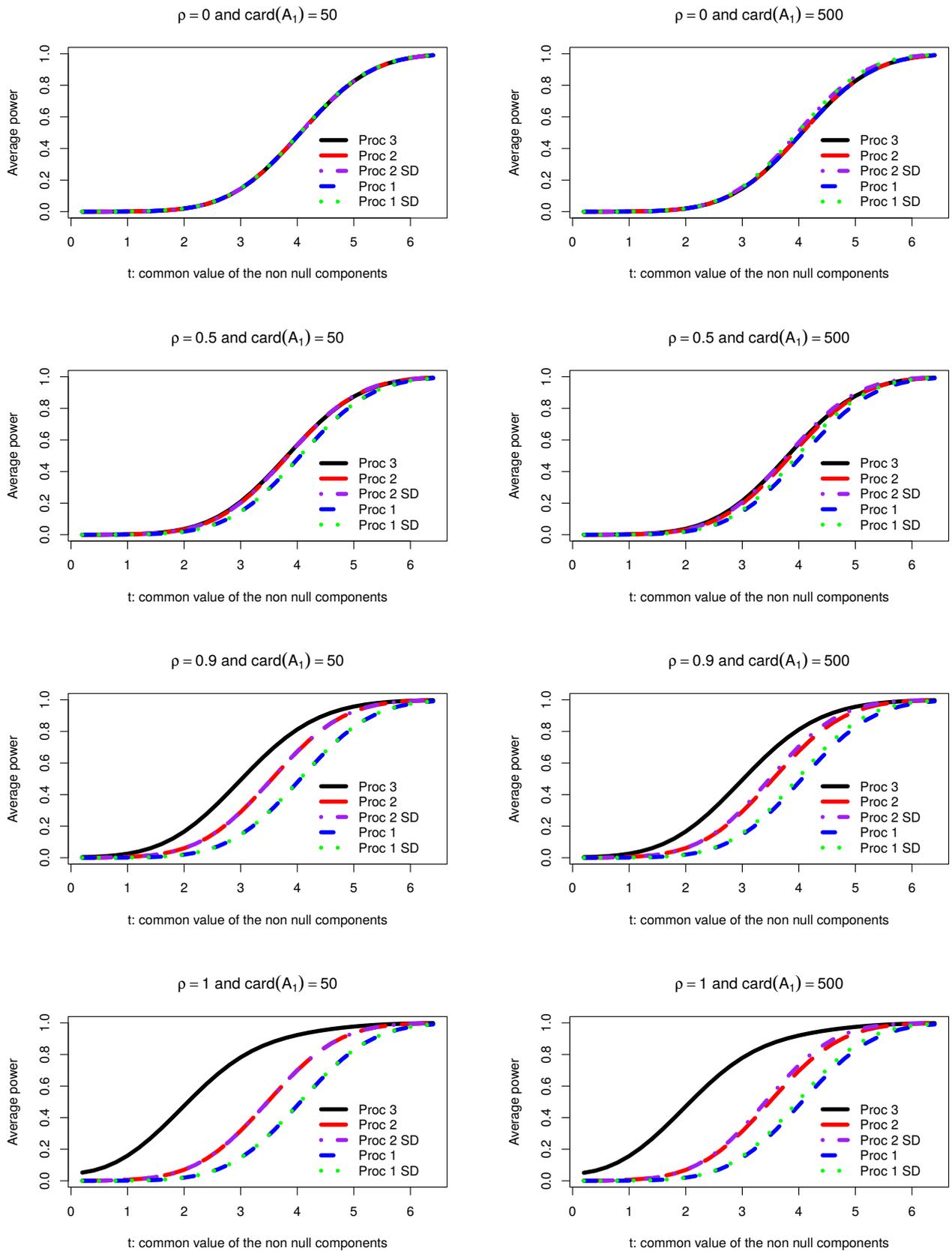


Figure 2: This figure provides the average power of procedures 1, 2, 1 SD, 2 SD and 3 (the average power of procedure 3 is reported in (4)). When $\rho = 0$, these five procedures have approximately the same power. When ρ increases, the difference between the average power of procedure 3 increases in comparison with the average power of the other procedures.

5 Application in metabolomics: detection of metabolites

Metabolomics is the science of detection of metabolites (small molecules) in biological mixtures (e.g. blood and urine). The most common technique for performing such characterization is proton nuclear magnetic resonance (NMR). Each metabolite generates a characteristic resonance signature in the NMR spectra with an intensity proportional to its concentration in the mixture. The number of peaks generated by a metabolite and their locations and ratio of heights are reproducible and uniquely determined: each metabolite has its own signature in the spectra. Each signature spectrum of each metabolite can be stored in a library that could contain hundreds of spectra. A major challenge in NMR analysis of metabolic profiles is automatic metabolite assignment from spectra. To identify metabolites, experts use spectra of pure metabolites and manually compare these spectra to the spectrum of the biological mixture under analysis. Such a method is time-consuming and requires domain-specific knowledge. Furthermore, complex biological mixtures can contain hundreds or thousands of metabolites, which can result in highly overlapping peaks. Figure 3 gives an example of an annotated spectrum of a mixture.

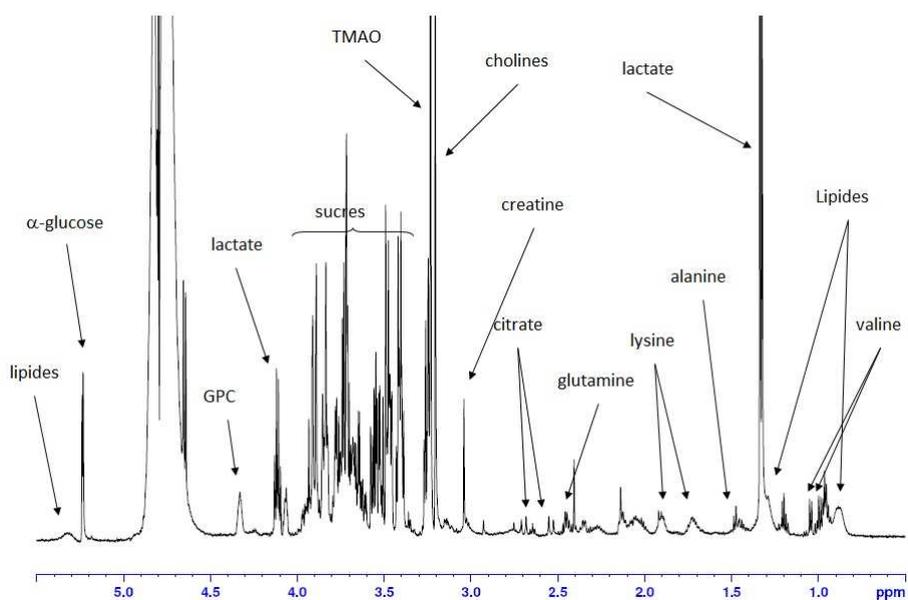


Figure 3: An annotated mixture spectrum with overlaps between peaks of lipides and valine and between peaks of glutamine and lysine.

Recently, automatic methods have been proposed, for example, Metabohunter [Tulpan et al., 2011], BATMAN [Astle et al., 2012, Hao et al., 2012], Bayesil [Ravanbakhsh et al., 2015] or the software Chenomx [Weljie et al., 2006]. Most of these methods are based on modelling using a Lorentzian shape and a Bayesian strategy. Nevertheless, most are time-consuming, and thus, cannot be applied to a large library of metabolites, and/or their statistical properties are not proven. Thus, the establishment of a gold-standard methodology with proven

statistical properties for identification of metabolites would be very helpful for the metabolomic community.

Because the number of tests is not very large (one can expect to analyse a mixture with about 200 metabolites), and as NMR experts want to recover all metabolites present in the mixture, but to prevent a false discovery, we developed a multiple testing procedure controlling the FWER.

5.1 Modelling

The spectrum of a metabolite (or a mixture) is a nonnegative function defined on a compact interval T . We assume that we have a library of spectra containing all $p = 36$ metabolites $\{f_i\}_{1 \leq i \leq p}$ (with $\int_T f_i(t)dt = 1$) that can be found in a mixture. This family of p spectra is assumed to be linearly independent. In the first approximation, the observed spectrum of the mixture Y can be modelled as a discretized noisy convex combination of the pure spectra:

$$Y_j = \left(\sum_{i=1}^p \beta_i^* f_i(t_j) \right) + \varepsilon_j \text{ with } 1 \leq j \leq n \text{ and } t_1 < \dots < t_n \text{ a subdivision of } T.$$

The random vector $(\varepsilon_1, \dots, \varepsilon_n)$ is a Gaussian vector $\mathcal{N}(\mathbf{0}, \Gamma)$, where Γ is a known and invertible covariance matrix. The covariance structure $(\varepsilon_1, \dots, \varepsilon_n)$ is described in [Tardivel et al., 2017, Tardivel, 2017].

5.2 Real dataset

The method for the detection of metabolites was tested on a known mixture. The NMR experts supplied us with a library of 36 spectra of pure metabolites and a mixture composed of these metabolites. The number of used metabolites and their proportions were unknown to us. The results are presented in Table 1.

Metabolites	Actual proportions	Significantly not null
Choline chloride	0.545	Yes
Creatinine	0.209	Yes
Benzoic acid	0.086	Yes
L-Proline	0.069	Yes
D-Glucose	0.060	Yes
L-Phenylalanine	0.029	Yes
30 other metabolites	0	No

Table 1: This table presents the results for the 36 metabolites of the library. The actual proportions of each metabolite are presented in the first column. For each metabolite, evidence against the nullity of the proportion is given in the second column.

The six metabolites present in the complex mixture are detected, including those with small proportions. There is no false discovery because any hypothesis associated with the 30 other metabolites was rejected. Because the whole procedure is fast, lasting only a few seconds, it could be easily applied to a library containing several hundred metabolites. We refer the interested reader to Tardivel et al. [2017] on this application to metabolomics, where our procedure is compared to existing ones on more complex datasets.

6 Conclusions

This study takes a new look at an old problem: the construction of multiple testing procedures derived from hyperrectangular confidence regions. Our purpose is to derive such a procedure based on hyperrectangular confidence regions having a minimal volume. These regions depend on an optimal threshold s^* , which is a solution of the constraint problem (2); we provide a solver giving a numerical solution to this problem. When $p \leq 1000$, the optimal threshold s^* (thus, the optimal hyperrectangular confidence region) is easily tractable. With respect to standard hypercube confidence regions, the gain in volume obtained with our method is huge. Based on simulations, we show that deriving a multiple testing procedure from a hyperrectangular region having a minimal volume is an intuitive way to increase the average power. Indeed, simulations show that our procedure is at least as powerful as the other procedures, and depending on the correlation matrix, our procedure can be much more powerful than the state-of-the-art procedures. However, it is still a challenge to provide a stepdown counterpart to our procedure.

7 Appendix

7.1 Proofs:

Proof of the proposition 1:

Proof of 1. We see that the volume cannot be minimal when $\|s\|_\infty$ is too large. Let $q_i > 0$ be the $1 - \alpha$ quantile of $|\zeta_i|$, let $q = \min\{q_1, \dots, q_p\} > 0$ and let us set $S := \{s \in \mathbb{R}^p \mid \mathbb{P}(|\zeta_1| \leq s_1, \dots, |\zeta_1| \leq s_p) = 1 - \alpha\}$. Whatever $s \in S$ the following inequality holds $\prod_{i=1}^p s_i \geq \|s\|_\infty q^{p-1}$, consequently, the function $s \in S \mapsto \prod_{i=1}^p s_i$ is coercive and continuous. Finally, by lemma 2 page 64 of Dentcheva [2006], S is closed; Thus, the minimum of the problem (2) is reached.

Proof of 2. It is straightforward that $(s_1^*/\sqrt{\Sigma_{11}}, \dots, s_p^*/\sqrt{\Sigma_{pp}})$ is a feasible point of the problem (3). Let $\tilde{s} \in [0, +\infty)^p$ be another feasible point of (3); thus, the following equivalence holds

$$\mathbb{P}\left(\frac{|\zeta_1|}{\sqrt{\Sigma_{11}}} \leq \tilde{s}_1, \dots, \frac{|\zeta_p|}{\sqrt{\Sigma_{pp}}} \leq \tilde{s}_p\right) = 1 - \alpha \Leftrightarrow \mathbb{P}\left(|\zeta_1| \leq \sqrt{\Sigma_{11}}\tilde{s}_1, \dots, |\zeta_p| \leq \sqrt{\Sigma_{pp}}\tilde{s}_p\right) = 1 - \alpha.$$

Consequently, $(\sqrt{\Sigma_{11}}\tilde{s}_1, \dots, \sqrt{\Sigma_{pp}}\tilde{s}_p)$ is a feasible point for the problem (1). Because s^* is a minimizer of the problem (2), one deduces that

$$\prod_{i=1}^p \sqrt{\Sigma_{ii}}\tilde{s}_i \geq \prod_{i=1}^p s_i^* \Leftrightarrow \prod_{i=1}^p \tilde{s}_i \geq \prod_{i=1}^p \frac{s_i^*}{\sqrt{\Sigma_{ii}}}.$$

□

Proof of the proposition 2:

To simplify the computation of the gradients, we consider the following problem, which has the same solution as (2)

$$\min f(s) = \sum_{i=1}^p \ln(s_i) \text{ subject to } F(s) = \mathbb{P}(|\zeta_1|/s_1 \leq 1, \dots, |\zeta_p|/s_p \leq 1) = 1 - \alpha.$$

As this problem reaches its minimum at s^* , $\nabla f(s^*)$ is collinear to $\nabla F(s^*)$. Let us set D the matrix $D = \text{diag}(s_1, \dots, s_p)$, we have the following expression for $F(s_1, \dots, s_p)$, namely,

$$F(s_1, \dots, s_p) = \int_{[-1,1]^p} R \exp\left(-\frac{1}{2}x^T DC^{-1}Dx\right) \det(D) dx = \int_{[-1,1]^p} R \exp\left(-\frac{1}{2}x^T DC^{-1}Dx + \ln(\det(D))\right) dx,$$

with $R = 1/((2\pi)^{p/2} \det(C)^{1/2})$. Next, the expression of the partial derivative

$$\frac{\partial}{\partial s_i} \left(-\frac{1}{2}x^T DC^{-1}Dx + \ln(\det(D)) \right) = \frac{1}{s_i} - \sum_{j=1}^p C_{i,j}^{-1} x_i x_j b_j,$$

implies that the gradient of F is equal to

$$\begin{aligned} \frac{\partial F}{\partial s_i}(s_1, \dots, s_p) &= \frac{1}{s_i} F(s_1, \dots, s_p) - R \sum_{j=1}^p \int_{[-1,1]^p} (C_{i,j}^{-1} x_i x_j b_j) \exp\left(-\frac{1}{2}x^T DC^{-1}Dx\right) \det(D) dx, \\ &= \frac{1 - \alpha}{s_i} - R \sum_{j=1}^p \int_{[-1,1]^p} (C_{i,j}^{-1} x_i x_j s_j) \exp\left(-\frac{1}{2}x^T DC^{-1}Dx\right) \det(D) dx. \end{aligned}$$

Thus, $\nabla F(s) = (1 - \alpha)\nabla f(s) + v(s)$, where $v(s) \in \mathbb{R}^p$ is the following vector

$$v(s) := \left(\sum_{j=1}^p C_{i,j}^{-1} \int_{[-1,1]^p} x_i x_j s_j^* R \exp\left(-\frac{1}{2}x^T DC^{-1}Dx\right) \det(D) dx \right)_{1 \leq i \leq p}.$$

Consequently, $\nabla f(s^*)$ and $\nabla F(s^*)$ are collinear if and only if $\nabla f(s^*)$ and $v(s^*)$ are collinear.

$$\begin{aligned} &\exists k \in \mathbb{R} \text{ such that } v(s^*) = k \nabla f(s^*), \\ \Leftrightarrow &\forall i \in \llbracket 1, p \rrbracket, \sum_{j=1}^p C_{i,j}^{-1} \int_{[-1,1]^p} x_i s_i^* x_j s_j^* R \exp\left(-\frac{1}{2}x^T DC^{-1}Dx\right) \det(D) dx = k, \\ \Leftrightarrow &\forall i \in \llbracket 1, p \rrbracket, \sum_{j=1}^p C_{i,j}^{-1} \int_{u \in \mathbb{R}^p} u_i u_j \frac{R}{1 - \alpha} \exp\left(-\frac{1}{2}u C^{-1} u\right) \mathbb{1}_{u \in s^*} du = \frac{k}{1 - \alpha}. \end{aligned} \quad (5)$$

The expression (5) is obtained *via* the change of variables $\forall i \in \llbracket 1, p \rrbracket, u_i = x_i s_i^*$. To conclude, one recognizes that

$$\int_{u \in \mathbb{R}^p} u_i u_j \frac{R}{1 - \alpha} \exp\left(-\frac{1}{2}u C^{-1} u\right) \mathbb{1}_{u \in s^*} du = \mathbb{E}\left(T_i^{s^*} T_j^{s^*}\right) = \text{cov}\left(T_i^{s^*}, T_j^{s^*}\right).$$

Thus, the diagonal coefficients of $C^{-1} \text{var}(T_{s^*})$ are equal to $k/(1 - \alpha)$. \square

7.2 Application of the generic stepdown method to procedure 3 does not control the FWER

Let us assume that the maximum likelihood estimator $\hat{\beta}^{\text{mle}}$ is distributed according to a $\mathcal{N}(\beta^*, C)$ distribution where C and β^* , as given hereafter,

$$C = (C_1|C_2|C_3) := \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \beta^* = (+\infty, 0, 0).$$

Let us apply the generic stepdown method to our procedure based on the optimal thresholds. The solution of the problem (2) with the matrix C and $\alpha = 0.05$ is $s_1^{(0)} = s_2^{(0)} = 2.10$ and $s_3^{(0)} = 2.43$. In the first step, the hypothesis \mathcal{H}_i is rejected if $|\hat{\beta}_i^{\text{mle}}| > s_i^{(0)}$ with $i \in \{1, 2, 3\}$. Obviously, because $\beta_1^* = +\infty$, the hypothesis \mathcal{H}_1 is rejected; let us assume that \mathcal{H}_2 and \mathcal{H}_3 are not rejected. Consequently, there is no false discovery in the first step; thus, $|\hat{\beta}_2^{\text{mle}}| \leq s_2^{(0)}$ and $|\hat{\beta}_3^{\text{mle}}| \leq s_3^{(0)}$. Let us set $\tilde{C} = (C_2|C_3) = Id_2$, the solution of the problem (2) with \tilde{C} is $s_2^{(1)} = s_3^{(1)} = 2.23$. In the second step, the hypothesis \mathcal{H}_i is rejected if $|\hat{\beta}_i^{\text{mle}}| > s_i^{(1)}$ with $i \in \{2, 3\}$. Let us assume that \mathcal{H}_2 and \mathcal{H}_3 are not rejected. Consequently, there is no false discovery in the second step; thus, $|\hat{\beta}_2^{\text{mle}}| \leq s_2^{(1)}$ and $|\hat{\beta}_3^{\text{mle}}| \leq s_3^{(1)}$. Finally, the probability of no false discovery is

$$\mathbb{P}(|\hat{\beta}_2^{\text{mle}}| \leq 2.10 \cap |\hat{\beta}_3^{\text{mle}}| \leq 2.23) < \mathbb{P}(|\hat{\beta}_2^{\text{mle}}| \leq 2.23 \cap |\hat{\beta}_3^{\text{mle}}| \leq 2.23) = 0.95.$$

Consequently, the FWER is not controlled at the significance level $\alpha = 0.05$.

Acknowledgements

The authors are grateful for the real data provided by the following metabolomicians from Toxalim: Cécile Canlet, Laurent Debrauwer, and Marie Tremblay-Franco and are grateful to Holger Rauhut for its careful reading. This work is part of the project GMO90+ supported by the grant CHORUS 2101240982 from the Ministry of Ecology, Sustainable Development and Energy in the national research program RiskOGM. Patrick Tardivel is partly supported by a PhD fellowship from GMO90+. We also received a grant for the project from the IDEX of Toulouse ‘Transversalité 2014’.

References

William Astle, Maria De Iorio, Sylvia Richardson, David Stephens, and Timothy Ebbels. A bayesian model of NMR spectra for the deconvolution and quantification of metabolites in complex biological mixtures. Journal of the American Statistical Association, 107(500):1259–1271, 2012.

- Yoav Benjamini, Vered Madar, and Philip B Stark. Simultaneous confidence intervals uniformly more likely to determine signs. *Biometrika*, 100(2):283–300, 2013.
- Darinka Dentcheva. Optimization models with probabilistic constraints. *Probabilistic and randomized methods for design under uncertainty*, pages 49–97, 2006.
- Sandrine Dudoit and Mark J Van Der Laan. *Multiple Testing Procedures with Applications to Genomics*. Springer, 2007.
- Olive Jean Dunn. Confidence intervals for the means of dependent, normally distributed variables. *Journal of the American Statistical Association*, 54(287):613–621, 1959.
- Olive Jean Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.
- Charles W Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121, 1955.
- Jie Hao, William Astle, Maria De Iorio, and Timothy MD Ebbels. BATMAN - an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a bayesian model. *Bioinformatics*, 28(15):2088–2090, 2012.
- Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.
- Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.
- Siamak Ravanbakhsh, Philip Liu, Trent C. Bjordahl, Rupasri Mandal, Jason R. Grant, Michael Wilson, Roman Eisner, Igor Sinelnikov, Xiaoyu Hu, Claudio Luchinat, Russell Greiner, and David S Wishart. Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE*, 10(5):e0124219, 2015.
- Joseph P Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.
- Thomas Royen. A simple proof of the gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East Journal of Theoretical Statistics*, 48(2):139–145, 2014.
- Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.
- Patrick Tardivel. *Représentation parcimonieuse et procédures de tests multiples: application à la métabolomique*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2017.

Patrick J. C. Tardivel, Cécile Canlet, Gaëlle Lefort, Marie Tremblay-Franco, Laurent Debrauwer, Didier Concordet, and Rémi Servien. ASICS: an automatic method for identification and quantification of metabolites in complex 1D ^1H NMR spectra. Metabolomics, 13(10):109, 2017.

Dan Tulpan, Serge Léger, Luc Belliveau, Adrian Culf, and Miroslava Čuperlović-Culf. Metabohunter: an automatic approach for identification of metabolites from ^1H -NMR spectra of complex mixtures. BMC Bioinformatics, 12(1):400, 2011.

Aalim M. Weljie, Jack Newton, Pascal Mercier, Erin Carlson, and Carolyn M. Slupsky. Targeted profiling: Quantitative analysis of ^1H -NMR metabolomics data. Analytical Chemistry, 78(13):4430–4442, 2006.