


# Hommel's procedure in linear time

Rosa J. Meijer<sup>1</sup> | Thijmen J. P. Krebs<sup>2</sup> | Jelle J. Goeman<sup>3</sup> 

<sup>1</sup>Statistics Netherlands, Postbus, HA Den Haag, The Netherlands

<sup>2</sup>Delft University of Technology, Mekelweg 4, CD Delft, The Netherlands

<sup>3</sup>Biomedical Data Sciences, Leiden University Medical Center, Postbus, RC Leiden, The Netherlands

## Correspondence

Jelle J. Goeman, Biomedical Data Sciences, Leiden University Medical Center, Postbus 9600, 2300 RC Leiden, The Netherlands  
Email: j.j.goeman@lumc.nl

## Funding information

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 639.072.412

## Abstract

Hommel's and Hochberg's procedures for familywise error control are both derived as shortcuts in a closed testing procedure with the Simes local test. Hommel's shortcut is exact but takes quadratic time in the number of hypotheses. Hochberg's shortcut takes only linear time after the  $P$ -values are sorted, but is conservative. In this paper, we present an exact shortcut in linear time on sorted  $P$ -values, combining the strengths of both procedures. The novel shortcut also applies to a robust variant of Hommel's procedure that does not require the assumption of the Simes inequality.

## KEYWORDS

close testing, convex hull, Hochberg, simes

## 1 | INTRODUCTION

The method of Hommel (1988) is a well-known multiple testing procedure that controls the familywise error rate (FWER), guaranteeing that with probability at least  $1 - \alpha$  no true null hypotheses are rejected. Hommel's method can be constructed using a combination of the closed testing procedure (Marcus, Peritz, & Gabriel, 1976) with local tests based on the inequality of Simes (1986). Hommel's procedure is uniformly more powerful than the methods of Bonferroni (Dunn, 1961), Holm (1979), and Hochberg (1988). A general and recent overview of  $P$ -value based procedures for FWER control can be found in Tamhane and Gou (2018).

Hommel's procedure is valid only if the Simes inequality can be assumed to hold for the  $P$ -values corresponding to true null hypotheses. The same condition is necessary for the validity of the procedure of Benjamini and Hochberg (1995) as a False Discovery Rate (FDR) controlling procedure. This condition has been extensively studied elsewhere (e.g. Benjamini and Yekutieli, 2001; Finner, Roters, & Strassburger, 2017; Rødland, 2006; Sarkar, 2008). A robust variant of the procedure, which is more conservative but does not require the assumption of the Simes inequality, was proposed by Hommel (1986).

In general, calculation time of a procedure based on closed testing is exponential in the number of hypotheses  $m$ . Faster algorithms, known as *shortcuts*, can be constructed for specific choices of local tests. Shortcuts are called exact if they always yield the same rejections as the full closed testing procedure, or conservative if they yield at most the same number of rejections, and sometimes fewer. For the case of local tests based on Simes, Hommel (1988) proposed an exact shortcut in quadratic time. Hochberg (1988) presented an alternative shortcut that takes linear time after the  $P$ -values are sorted (which takes linearithmic time). Hochberg's shortcut, however, is conservative. A hybrid Hochberg–Hommel procedure (Gou, Tamhane, Xi, & Rom, 2014) exists, that can be more powerful than Hommel's procedure, but only approximate algorithms are available.

In this paper, we present a novel exact shortcut for closed testing with local tests based on Simes local tests. The new shortcut is exact like Hommel's procedure, but takes only linear time on the sorted  $P$ -values like Hochberg's. It allows computationally efficient scaling of Simes-based multiple testing procedures to large multiple testing problems without the power loss incurred by switching to Hochberg's method. The new shortcut also generalizes to the robust variant of Hommel's method (Hommel, 1986) that does not assume the Simes inequality.

The structure of the paper is as follows. We start by introducing closed testing and the Simes local test. From a result by Hommel, we introduce the crucial function  $h(\alpha)$  and propose a novel algorithm to calculate it in linear time after sorting the  $P$ -values. Next, we explain how to calculate all adjusted  $P$ -values from  $h(\alpha)$  in linear time. We finish with a short simulation that compares runtimes in practice, and a data analysis example that compares adjusted  $P$ -values between Hochberg's and Hommel's procedure. The algorithms described in this paper have been implemented in the R-package `hommel`, available on CRAN.

## 2 | CLOSED TESTING, SIMES, HOMMEL, AND HOCHBERG

Suppose we have  $m$  hypotheses  $H_1, \dots, H_m$  that we are interested in testing. Some of the hypotheses are true, while others are false. Denote by  $T \subseteq \{1, \dots, m\}$  the index set of true hypotheses. For each hypothesis we suppose that we have corresponding (unadjusted)  $P$ -values  $p_1, \dots, p_m$ , one for each hypothesis  $H_1, \dots, H_m$ . We assume that the  $P$ -values will be sorted, so that  $p_1 \leq \dots \leq p_m$ . Sorting the  $P$ -values takes linearithmic time (i.e.  $O(m \log m)$  time in big  $O$  notation).

The aim of a classical multiple testing procedure is to find an index set  $R_\alpha$  of hypotheses to reject, which is as large as possible while still controlling FWER at level  $\alpha$ . This requires that

$$P(T \cap R_\alpha = \emptyset) \geq 1 - \alpha,$$

that is, with probability at least  $1 - \alpha$  there are no type I errors. It is often convenient to describe  $R_\alpha$  as a function of  $\alpha$  through adjusted  $P$ -values  $\tilde{p}_1, \dots, \tilde{p}_m$ , defined as follows:

$$\tilde{p}_i = \min\{0 \leq \alpha \leq 1 : i \in R_\alpha\}.$$

Generally,  $R_\alpha$  is monotone in  $\alpha$  with  $R_1 = \{1, \dots, m\}$  so that all  $\tilde{p}_i$  are defined and we have  $R_\alpha = \{1 \leq i \leq m : \tilde{p}_i \leq \alpha\}$ .

Like Hommel and Hochberg, we consider sets  $R_\alpha$  arising from the combination of closed testing with Simes tests. The closed testing procedure (Marcus et al., 1976) augments the collection of hypotheses with all possible intersection hypotheses  $H_I = \bigcap_{i \in I} H_i$ , with  $I \subseteq \{1, \dots, m\}$ . An intersection hypothesis  $H_I$  is true if and only if  $H_i$  is true for all  $i \in I$ . Note that  $H_i = H_{\{i\}}$ , so that all original hypotheses, known as *elementary hypotheses*, are also intersection hypotheses. Next, all intersection hypotheses are tested with a valid  $\alpha$ -level test, the *local test*. The set  $R_\alpha$  is now defined as the set of all  $i \in \{1, \dots, m\}$  for which  $H_I$  is rejected by the local test for all  $I \ni i$ . For FWER to be controlled it is sufficient that the local test rejects  $H_T$ , the intersection of all true hypotheses, with probability at most  $\alpha$  (Goeman & Solari, 2011).

The Simes local test rejects an intersection hypothesis  $H_I$  if and only if there is at least one  $i \in \{1, \dots, |I|\}$  for which

$$p_{(i:I)} \leq \frac{i\alpha}{|I|}, \quad (1)$$

where for any  $I \subseteq \{1, \dots, m\}$  and  $i \in \{1, \dots, |I|\}$ , we define  $p_{(i:I)}$  as the  $i$ -th smallest  $P$ -value among the sequence  $(p_j)_{j \in I}$ .

For the validity of the resulting closed testing procedure we must assume that the Simes test rejects  $H_T$  with probability at most  $\alpha$ . There is extensive and growing literature on the conditions under which this assumption holds, which we will not revisit here (Benjamini & Yekutieli, 2001; Finner et al., 2017; Rødland, 2006; Sarkar, 2008). The validity of the Simes test on  $H_T$  is also necessary for the validity of the procedure of Benjamini and Hochberg (1995) as an FDR-controlling procedure.

A robust variant of the Simes test due to Hommel (1983) rejects an intersection hypothesis  $H_I$  if and only if there is at least one  $i \in \{1, \dots, |I|\}$  for which

$$p_{(i:I)} \leq \frac{i\alpha}{|I| \sum_{j=1}^{|I|} j^{-1}}. \quad (2)$$

This robust local test is more conservative but makes no assumption on the joint distribution of the  $P$ -values: it is valid whenever the distribution of  $P$ -values from true null hypotheses is uniform or stochastically larger than that. Its use as a local test in combination with closed testing was proposed by Hommel (1986).

In our discussion below, we will formulate results simultaneously for the tests (1) and (2) by treating them as special cases of a general local test that rejects  $H_I$  if and only if there is at least one  $i \in \{1, \dots, |I|\}$  for which

$$s_{|I|} p_{(i:I)} \leq i\alpha. \quad (3)$$

The interesting cases will be  $s_k = k$  (Simes) and  $s_k = k \sum_{j=1}^k j^{-1}$  (robust variant), but in the rest of this paper we will only use that  $s_0 = 0$  and that  $s_k$  is weakly increasing (i.e. nondecreasing) in  $k$ . We also assume that all  $s_k$ ,  $k = 1, \dots, m$  can be calculated together in at most linear time. Trivially, we reject all hypotheses when  $\alpha = 1$ . Note that we do not make any claims for validity or usefulness of the general test (3) aside from the two particular cases introduced above. For a general family with proven properties as a local test that contains the robust test as a special case, see Blanchard and Roquain (2008). That general family does not admit an exact linear shortcut, however.

Naive application of the closed testing procedure requires  $2^m - 1$  local tests to be performed, which severely limits the usefulness of the procedure in large problems. For this reason *shortcuts* have been developed for specific local tests. For the Simes local test Hommel (1988) proved that  $i \in R_\alpha$  if and only if

$$h(\alpha)p_i \leq \alpha, \quad (4)$$

where  $h(\alpha)$  is defined in Section 3 below. Note that Hommel's rule is analogous to Bonferroni, except that the Bonferroni factor  $m$  is replaced by the random variable  $h(\alpha) \leq m$ . Based on this result Hommel formulated a quadratic time algorithm for calculating all adjusted  $P$ -values.

To obtain an alternative shortcut for the Simes case, Hochberg (1988) proved that  $R'_\alpha \subseteq R_\alpha$ , where  $i \in R'_\alpha$  if and only if there is a  $j \geq i$  such that

$$(m - j + 1)p_j \leq \alpha.$$

This result leads to an easy linear time shortcut if the  $P$ -values are sorted, or a linearithmic time one if the sorting is taken into account. Hochberg's method is conservative when  $m > 2$ , sacrificing power for computational efficiency (Hommel, 1989). It is easy to find instances for which  $R'_\alpha \subset R_\alpha$ . For example when  $m = 4$ ,  $\alpha = 0.05$ ,  $p_1 = 0.02$ ,  $p_2 = 0.02$ ,  $p_3 = 0.03$ , and  $p_4 = 0.90$ , Hommel's method rejects two hypotheses, while Hochberg's method rejects none.

### 3 | THE FUNCTION $h(\alpha)$

To construct the new shortcut we start from Hommel's result (4), generalizing it to the general local test (3). The quantity in Hommel's procedure that is most costly to calculate is the function  $h(\alpha)$ . In this section, we will study the properties of this function, before we show how a representation can be calculated for all  $\alpha$  in linear time.

We define  $h(\alpha)$  as the largest size  $|I|$  of an intersection hypothesis  $H_I$  that cannot be rejected by the closed testing procedure at level  $\alpha$ . Equivalently,  $h(\alpha)$  is therefore also the largest size  $|I|$  of an intersection hypothesis  $H_I$  that cannot be rejected by the local test at level  $\alpha$ . If any hypothesis of size  $i$  is not rejected, then certainly the “worst case” hypothesis  $H_{K_i}$  with  $|K_i| = i$  given by

$$K_i = \{m - i + 1, \dots, m\}$$

is not rejected. Applying (3) it follows that

$$h(\alpha) = \max \{i \in \{0, \dots, m\} : s_i p_{m-i+j} > j\alpha, \text{ for } j = 1, \dots, i\}.$$

The importance of  $h(\alpha)$  is clear from the following lemma.

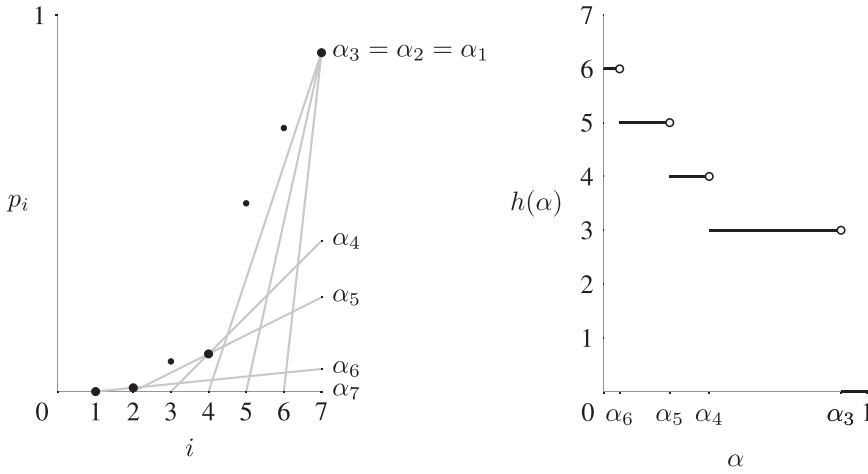
**Lemma 1.**  $H_i$  is rejected by the closed testing procedure if and only if  $s_{h(\alpha)}p_i \leq \alpha$ .

*Proof.* Write  $h = h(\alpha)$ . First, assume that  $H_i$  is rejected by the closed testing procedure, and without loss of generality that  $h > 0$ . Then  $H_I$  is rejected for all  $I \ni i$ . Since  $H_{K_h}$  is not rejected, we have  $i \notin K_h$ . Since  $H_{K_{h-1} \cup \{i\}}$  is rejected, but  $H_{K_h}$  is not,  $s_h p_i \leq \alpha$  follows from comparing their local tests.

Next, assume  $s_h p_i \leq \alpha$  and take  $I \ni i$ . Now either  $|I| > h$ , so  $H_I$  is rejected by the local test, or  $|I| \leq h$ , and

$$s_{|I|} p_{(1:I)} \leq s_{|I|} p_i \leq s_h p_i \leq \alpha,$$

so  $H_I$  is rejected by the local test. Consequently,  $H_I$  is rejected by the local test for all  $I \ni i$ , so  $H_i$  is rejected by the closed testing procedure.  $\square$



**FIGURE 1** On the left-hand side an example of the construction of  $\alpha_i$ ,  $i = 1, \dots, m$ , for the Simes local test with  $m = 7$  and  $p_1 = 0$ ,  $p_2 = 0.01$ ,  $p_3 = 0.08$ ,  $p_4 = 0.1$ ,  $p_5 = 0.5$ ,  $p_6 = 0.7$ ,  $p_7 = 0.9$ . The dots are the  $P$ -values plotted against their rank, with the bigger dots indicating the support of the lower convex hull (Section 4). The value of  $\alpha_i = \alpha_i^*$  is the largest  $\alpha$  such that the line from  $(m - i, 0)$  to  $(m, \alpha)$  has no points below the line. For the given  $P$ -values we find  $\alpha_1 = \alpha_2 = \alpha_3 = p_7 = 0.9$ ,  $\alpha_4 = 4p_4 = 0.4$ ,  $\alpha_5 = 5p_4/2 = 0.25$ ,  $\alpha_6 = 6p_2 = 0.06$ ,  $\alpha_7 = 7p_1 = 0$ . The right-hand side is the resulting step function  $h(\alpha)$

The function  $h(\alpha)$  is a right-continuous step function on its domain  $[0, 1]$ , since it is weakly decreasing and takes integer values in  $\{0, \dots, m\}$ . To describe  $h(\alpha)$  it suffices, therefore, to find the jumps of the function. We have  $h(1) = 0$ , and, if  $s_1 > 0$ ,  $h(0)$  is the number of nonzero  $P$ -values. Let

$$\alpha_i = \min\{0 \leq \alpha \leq 1 : h(\alpha) < i\}.$$

By definition of  $h(\alpha)$ ,  $\alpha_i$  for  $1 \leq i \leq m$  is the lowest  $\alpha$ -level at which the hypothesis  $H_{K_j}$ , the “worst case hypothesis of size  $j$ ,” can still be rejected by the local test procedure for all  $j \geq i$ . Therefore, for  $i = 1, \dots, m$ , we have

$$\alpha_i = \max_{i \leq j \leq m} \alpha_j^*, \quad (5)$$

where  $\alpha_j^*$  is the lowest  $\alpha$ -level at which the hypothesis  $H_{K_j}$  can be rejected by the local test. By (3) we have, after changing variables,

$$\alpha_i^* = \min_{j \in \{m-i+1, \dots, m\}} \frac{s_i p_j}{j - m + i}. \quad (6)$$

For  $i > m$  we have  $\alpha_i = \alpha_i^* = 0$ . Tied values of  $\alpha_i$  indicate jumps of  $h(\alpha)$  larger than 1. Since we may trivially reject all hypotheses if  $\alpha = 1$  we may set all values of  $\alpha_i^*$  that exceed 1 to 1.

In the case of the Simes local test, things simplify slightly, since in that case  $\alpha_i = \alpha_i^*$  for all  $i$ . Namely, for any  $k \geq i$ , we have

$$\alpha_k^* \leq \min_{j \in \{m-i+1, \dots, m\}} \frac{kp_j}{j - m + k} \leq \alpha_i^*,$$

because  $k/(j - m + k) \leq i/(j - m + i)$  for each  $m - i + 1 \leq j \leq m$ . We also remark that for Simes local tests we can only have  $\alpha_i = \alpha_j$  for  $i \neq j$  if  $\alpha_i = \alpha_j = 0$  or if  $\alpha_i = \alpha_j = \alpha_m = p_m$ . Therefore the jumps of  $h(\alpha)$  are all of size 1 except possibly the first and the last one. This is intuitive from Figure 1, where we illustrate the function  $h(\alpha)$  and its jumps for the case of Simes local tests.

## 4 | A LINEAR TIME ALGORITHM FOR $h(\alpha)$

As seen in Figure 1, for the case of the Simes local test the value of  $\alpha_i^*$  is the largest  $\alpha$  such that the line from  $(m - i, 0)$  to  $(m, \alpha)$  is never above any of the points  $(i, p_i)$ . A similar observation holds for the general test. This in turn suggests that not all  $(i, p_i)$  are important for the calculation of  $h(\alpha)$ , but only the points that support the lower convex hull, that is the part of the convex hull at or below the line through  $(0, 0)$  and  $(m, p_m)$ . This observation can be used to calculate  $h(\alpha)$  on the sorted  $P$ -values in linear time, as we will show in this section.

It will be convenient to add  $p_0 = 0$ . Let  $0 = c_0 < \dots < c_r = m$  be indices such that  $(c_0, p_0), \dots, (c_r, p_{c_r})$  form the lower convex hull of the points  $(0, p_0), \dots, (m, p_m)$ . If the  $P$ -values are sorted, the values  $c_0, \dots, c_r$  can be found in linear time using the algorithms of Andrew (1979) or Fortune (1989). The latter is preferred, since it is numerically more stable. It is given below

as Algorithm 1. The algorithm is easily understood from the example of Figure 1 for the Simes case that  $s_k = k$  for all  $k$ . It builds the lower hull from left to right. At step  $i$  of the algorithm, the point  $(i, p_i)$  is added to the hull. Next, all previously added points are checked, in a last-in-first-out manner, and discarded if above the current hull. The if-statement is the part added due to Fortune (1989). It checks that a point  $(i, p_i)$  is indeed a candidate for the lower, rather than the upper hull, before adding it.

### Algorithm 1: Fortune's algorithm for the lower convex hull of the points $(0, p_0), (1, p_1), \dots, (m, p_m)$

---

**Require:**  $0 = p_0 \leq p_1 \leq \dots \leq p_m \leq 1$

**function** LOWERCONVEXHULL  $p_0, \dots, p_m$

$p_{-1} \leftarrow 1$

$c_{-1}, c_0 \leftarrow -1, 0$

$r \leftarrow 0$

**for**  $i \leftarrow 1$  **to**  $m$  **do**

**if**  $i = m$  **or**  $mp_i < ip_m$  **then**

**while**  $(i - c_r)(p_{c_r} - p_{c_{r-1}}) \geq (c_r - c_{r-1})(p_i - p_{c_r})$  **do**

$r \leftarrow r - 1$

$r \leftarrow r + 1$

$c_r \leftarrow i$

**return**  $c_0, \dots, c_r$

---

We can use the lower convex hull to find the  $\alpha_i^*$  by the following lemma.

**Lemma 2.** *The minimum in (6) is attained at  $c_{\max\{1 \leq k \leq r: D_k \geq 0\}}$ , where*

$$D_k = p_{c_{k-1}}(c_k - m + i) - p_{c_k}(c_{k-1} - m + i). \quad (7)$$

*Proof.* Note that the maximum  $\tilde{k}$  exists, and  $c_{\tilde{k}} > m - i$ . Indeed, there is some  $1 \leq k \leq r$  such that  $c_{k-1} \leq m - i < c_k$ , so  $D_k \geq 0$ . It suffices to show that for all  $m - i < j \leq m$  that

$$\frac{p_j}{j - m + i} \geq \frac{p_{c_{\tilde{k}}}}{c_{\tilde{k}} - m + i}. \quad (8)$$

Since  $(j, p_j)$  lies on or above any line containing a lower hull edge, we have for  $1 \leq k \leq r$  that

$$p_j \geq \frac{p_{c_{k-1}}(c_k - j) + p_{c_k}(j - c_{k-1})}{c_k - c_{k-1}} = \frac{D_k + (p_{c_k} - p_{c_{k-1}})(j - m + i)}{c_k - c_{k-1}},$$

so

$$\frac{p_j}{j - m + i} \geq \frac{D_k / (j - m + i) + p_{c_k} - p_{c_{k-1}}}{c_k - c_{k-1}}.$$

If  $j \leq c_{\tilde{k}}$ , we take  $k = \tilde{k}$  and see that the right-hand side is minimized at  $j = c_{\tilde{k}}$  as  $D_k \geq 0$ , so that (8) follows. If  $j \geq c_{\tilde{k}}$  and  $\tilde{k} < r$ , we take  $k = \tilde{k} + 1$  and see that the right-hand side is minimized at  $j = c_{\tilde{k}}$  as  $D_k \leq 0$ , so that (8) follows.  $\square$

The result of Lemma 2 can be used with the following algorithm, which additionally uses the observation that if  $\tilde{k}_i$  is the maximum for  $\alpha_i^*$ , and  $\tilde{k}_{i+1}$  is the maximum for  $\alpha_{i+1}^*$ , then  $\tilde{k}_i \geq \tilde{k}_{i+1}$ , since  $D_{\tilde{k}_{i+1}} \geq 0$  implies that  $D_{\tilde{k}_i} \geq 0$ . Note also that  $D_1 \geq 0$ , so that  $k$  remains positive.

## Algorithm 2: Compute the lowest $\alpha$ -levels on which the hypotheses $H_{K_1}, \dots, H_{K_m}$ are rejected by the closed testing procedure

---

**Require:**  $0 \leq p_1 \leq \dots \leq p_m \leq 1, 0 = s_0 \leq s_1 \leq \dots \leq s_m$

**function** LOWESTALPHALEVELS( $p_1, \dots, p_m, s_0, \dots, s_m$ )

$p_0 \leftarrow 0$

$c_0, \dots, c_k \leftarrow \text{LOWERCONVEXHULL}(p_0, p_1, \dots, p_m)$

**for**  $i \leftarrow 1$  **to**  $m$  **do** ▷ Use Lemma 2

**while**  $p_{c_{k-1}}(c_k - m + i) < p_{c_k}(c_{k-1} - m + i)$  **do** ▷  $D_k < 0$

$k \leftarrow k - 1$

$\alpha_i^* \leftarrow s_i p_{c_k} / (c_k - m + i)$

$\alpha_{m+1} \leftarrow 0$  ▷ Use equation (6)

**for**  $i \leftarrow m$  **to**  $1$  **do** ▷ With Simes local tests simply  $\alpha_i = \alpha_i^*$

$\alpha_i \leftarrow \max(\alpha_i^*, \alpha_{i+1})$

**return**  $\alpha_1, \dots, \alpha_{m+1}$

---

Finally, we can obtain  $\alpha_1, \dots, \alpha_m$  by (5) in linear time. Note that this last step may be omitted in the case of the Simes local test, since in that case  $\alpha_i^* = \alpha_i$  for all  $i$ .

## 5 | ADJUSTED $P$ -VALUES

From  $h(\alpha)$  we can subsequently calculate the adjusted  $P$ -value  $\tilde{p}_i$  for each hypothesis  $H_i$  in Hommel's procedure by finding the minimum  $\alpha$  for which  $s_{h(\alpha)} p_i \leq \alpha$ . We can, however, find all the adjusted  $P$ -values in linear time when we exploit the following lemma. The new algorithm can be seen as an update to the quadratic time algorithm of Wright (1992). Note that  $\alpha_{m+1} = 0$  by definition.

**Lemma 3.** *The adjusted  $P$ -value of  $H_i$  is given by*

$$\tilde{p}_i = \min(s_j p_i, \alpha_j),$$

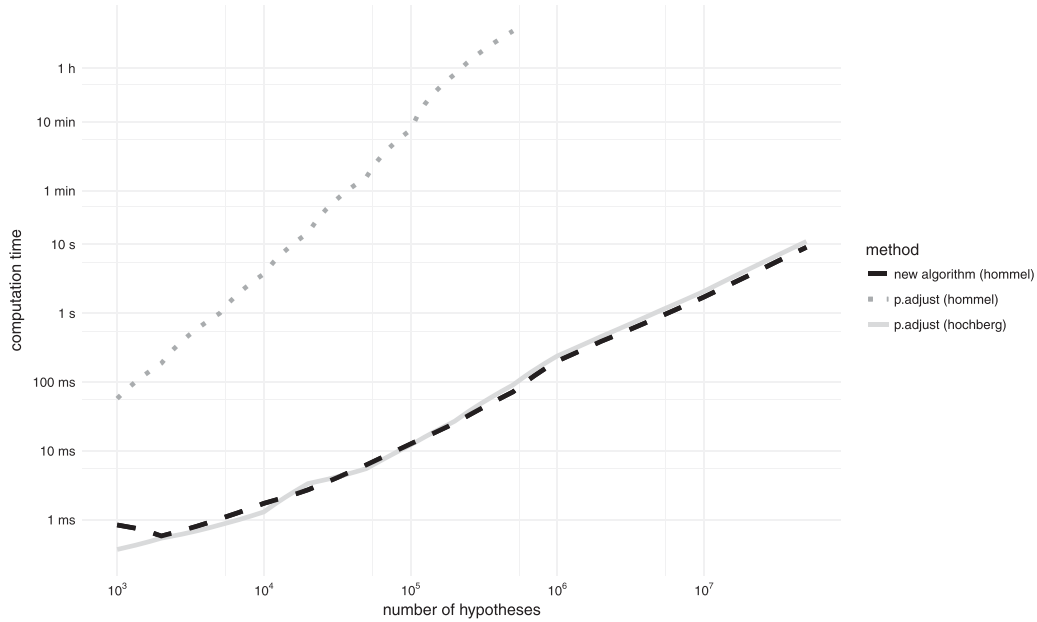
where

$$\tilde{j} = \max\{1 \leq j \leq m+1 : s_{j-1} p_i \leq \alpha_j\}. \quad (9)$$

*Proof.* Let  $\alpha \geq \tilde{p}_i$ . Then  $\alpha \geq \alpha_j$  or  $\alpha \geq s_j p_i$ . If  $\alpha \geq \alpha_j$ , then  $h(\alpha) < j$ . By definition of  $\tilde{j}$  we have  $s_{h(\alpha)} p_i \leq s_{j-1} p_i \leq \alpha_j \leq \alpha$ , which implies by Lemma 1 that  $H_i$  is rejected by the closed testing procedure. If  $\alpha \geq s_j p_i$ , then by definition of  $\tilde{j}$  if  $\tilde{j} \leq m$  and trivially if  $\tilde{j} = m+1$ , we have  $\alpha \geq s_{\tilde{j}} p_i \geq \alpha_{\tilde{j}+1}$ . Therefore  $h(\alpha) \leq \tilde{j}$ , so that  $s_{h(\alpha)} p_i \leq s_{\tilde{j}} p_i \leq \alpha$ , so  $H_i$  is rejected by the closed testing procedure by Lemma 1.

Let  $\alpha < \tilde{p}_i$ . Then  $\alpha < s_j p_i$  and  $\alpha < \alpha_j$ . The latter implies that  $h(\alpha) \geq j$ , so that  $s_{h(\alpha)} p_i \geq s_j p_i > \alpha$ . This implies by Lemma 1 that  $H_i$  is not rejected by the closed testing procedure.  $\square$

We can use the procedure below to find  $\tilde{p}_1, \dots, \tilde{p}_m$ . To see that the procedure is valid, let  $\tilde{j}_i$  be the value for the maximum for the calculation of  $\tilde{p}_i$ . Note that for each  $i$  the procedure finds the largest value of  $1 \leq j \leq \tilde{j}_{i-1}$  such that  $s_{j-1} p_i \leq \alpha_j$ . Since  $p_i \geq p_{i-1}$  we have  $\tilde{j}_i \leq \tilde{j}_{i-1}$ , so this is equivalent to (9). It is obvious that Algorithm 3 takes linear time in  $m$ , and that consequently all adjusted  $P$ -values can be calculated in linear time if  $h(\alpha)$  is known.



**FIGURE 2** Computation time for the new algorithm as implemented in the `hommel` package compared with Hommel's and Hochberg's methods as implemented in `p.adjust`

**Algorithm 3: Compute adjusted  $P$ -values for the hypotheses  $H_1, \dots, H_m$  under the local test (3) described by  $s_0, \dots, s_m$**

**Require:**  $0 \leq p_1 \leq \dots \leq p_m \leq 1, 0 = s_0 \leq s_1 \leq \dots \leq s_m$

```

function ADJUSTEDPVALUES( $p_1, \dots, p_m, s_0, \dots, s_m$ )
   $\alpha_1, \dots, \alpha_{m+1} \leftarrow \text{LOWESTALPHALEVELS}(p_1, \dots, p_m, s_0, \dots, s_m)$ 
   $j \leftarrow m + 1$  ▷ Use Lemma 3
   $s_j \leftarrow 0$ 
  for  $i \leftarrow 1$  to  $m$  do
    while  $s_{j-1}p_i > \alpha_j$  do
       $j \leftarrow j - 1$ 
     $\tilde{p}_i \leftarrow \min(s_j p_i, \alpha_j)$ 
  return  $\tilde{p}_1, \dots, \tilde{p}_m$ 

```

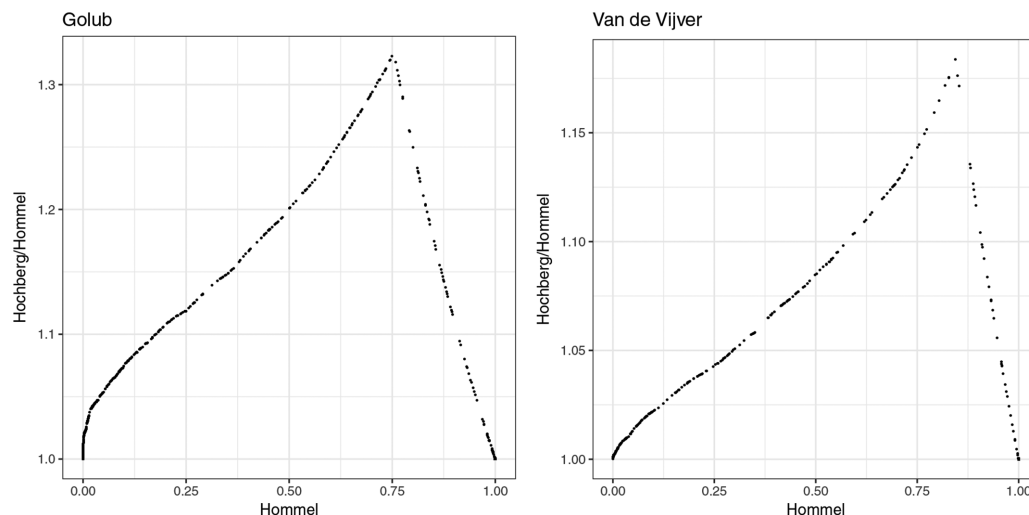
## 6 | IMPLEMENTATION

The algorithms in this paper have been implemented in R in the package `hommel` that is available on CRAN. We compare computation time in practice between the new algorithm and the current implementation of Hommel's and Hochberg's algorithms in R's function `p.adjust`.

We sampled  $m$   $P$ -values independently from a squared uniform distribution, varying  $m$  from  $10^3$  to  $5 \times 10^7$ . Computation times were evaluated from the `p.adjust`-function (package `stats`) for both the Hommel and Hochberg methods, and for the new algorithm as implemented in the `hommel` package. The results are given in Figure 2. Computation times for Hommel's method with `p.adjust` were not calculated for  $m$  above  $5 \times 10^5$  since at this value of  $m$  the calculation already took more than 3 hours (against 0.1 second for the new algorithm). Computation times less than 10 minutes have been averaged over several runs. The exact R code used to generate the figure is included in the Supplementary Information.

The computational gain of the new shortcut relative to Hommel's method is clear. We see that there is no apparent advantage for Hochberg's method over the new exact algorithm in terms of computation time, because in both algorithms the dominant factor in the computation is the sorting of the  $P$ -values.





**FIGURE 3** Ratio of adjusted  $P$ -values of Hochberg's and Hommel's method as a function of the adjusted  $P$ -value of Hommel's method for two gene expression data sets. Left-hand side: Golub ( $m = 7,129$ ); right-hand side: Van de Vijver ( $m = 24,481$ )

## 7 | APPLICATION

To check the gain of the new algorithm over Hochberg's method in large data sets we applied the new algorithm on two classical gene expression data. As a first data set we used the data of Golub et al. (1999) with  $m = 7,129$  hypotheses based on the `golubEsets` R package (Golub, 2017). The data were normalized by `vsr` (Huber, Von Heydebreck, Sültmann, Poustka, & Vingron, 2002) and raw  $P$ -values for association of gene expression with AML versus ALL were calculated using `limma` (Smyth, 2004). The second dataset was by Van de Vijver et al. (2002) with  $m = 24,481$  hypotheses based on the `breastCancerNKI` R package (Schroeder et al., 2011). The outcome was distant metastasis-free survival, and raw  $P$ -values were calculated using score tests in a Cox proportional hazards model. The exact R code used to generate the figures is included in the Supplementary Material.

Figure 3 displays the relative loss in adjusted  $P$ -value of Hochberg's procedure as a function of the Hommel adjusted  $P$ -value. We see that the gain of Hommel's procedure relative to Hochberg's is small. It is most appreciable in the middle range. Around the relevant 0.05 level Hochberg adjusted  $P$ -values are less than a factor 1.05 too large. The gain of Hommel versus Hochberg in the Golub data is larger because there is more signal in that dataset. The peak and the diagonal line shows that that Hochberg's method sets many more adjusted  $P$ -values approximately to 1 than Hommel's method. The gain of Hommel's method relative to Hochberg is small, but it is free. There is no loss in computation time using Hommel's method with the new algorithm. Computation times for the adjusted  $P$ -values were less than 0.02 seconds on a standard PC for both methods in both datasets. In contrast, the quadratic time algorithm in `p.adjust` took 3.5 seconds in the Golub data and 38 seconds in the Van de Vijver data.

## 8 | DISCUSSION

We have presented a new shortcut that calculates adjusted  $P$ -values of elementary hypotheses for closed testing with Simes local tests on sorted  $P$ -values in linear time. The new method combines the best of Hommel's and Hochberg's methods: it has the computational speed of Hochberg's method and the power of Hommel's method.

The computational gain of the novel shortcut in this paper is important when many hypotheses are tested and FWER control is desired. Although FDR control is preferred in some application areas with many hypotheses (e.g. transcriptomics), FWER control remains the standard in other areas where more rigorous error control is desired, for example genome-wide association studies (Sham & Purcell, 2014) and neuroimaging (Eklund, Nichols, & Knutsson, 2016). In the end, the choice for FDR or FWER should not be made on the basis of the size of the multiple testing problem, but on the desired use and reliability of the discoveries. FWER control, unlike FDR control, for example, has the important property that error control remains guaranteed for arbitrary subsets of the discoveries (Finner & Roters, 2001; Goeman & Solari, 2014).



This paper focuses on elementary hypotheses only, while Hommel's procedure also gives additional nonconsonant rejections. Such nonconsonant rejections can be used for calculating simultaneous confidence bounds for the false discovery proportion of all subsets of the hypotheses (Goeman & Solari, 2011). The results of this paper may also be instrumental for obtaining such bounds more quickly (Goeman, Meijer, Krebs, & Solari, 2017).

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## ORCID

Jelle J. Goeman  <http://orcid.org/0000-0003-4283-0259>

## REFERENCES

- Andrew, A. M. (1979). Another efficient algorithm for convex hulls in two dimensions. *Information Processing Letters*, 9(5), 216–219.
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165–1188.
- Blanchard, G. & Roquain, E. (2008). Two simple sufficient conditions for fdr control. *Electronic Journal of Statistics*, 2, 963–992.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293), 52–64.
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28), 7900–7905.
- Finner, H. & Roters, M. (2001). On the false discovery rate and expected type I errors. *Biometrical Journal*, 43(8), 985–1005.
- Finner, H., Roters, M., & Strassburger, K. (2017). On the Simes test under dependence. *Statistical Papers*, 58(3), 775–789.
- Fortune, S. (1989). Stable maintenance of point set triangulations in two dimensions. In *Foundations of Computer Science, 1989., 30th Annual Symposium on* (pp. 494–499). IEEE, Piscataway, NJ.
- Goeman, J., Meijer, R., Krebs, T., & Solari, A. (2017). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *arXiv preprint arXiv:1611.06739*.
- Goeman, J. J. & Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, 26(4), 584–597.
- Goeman, J. J. & Solari, A. (2014). Multiple hypothesis testing in genomics. *Statistics in Medicine*, 33(11), 1946–1978.
- Golub, T. (2017). golubEsets: exprSets for golub leukemia data. R package version 1.18.0.
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., ... Lander, E. S. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439), 531–537.
- Gou, J., Tamhane, A. C., Xi, D., & Rom, D. (2014). A class of improved hybrid Hochberg–Hommel type step-up multiple test procedures. *Biometrika*, 101(4), 899–911.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4), 800–802.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hommel, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal*, 25, 423–430.
- Hommel, G. (1986). Multiple test procedures for arbitrary dependence structures. *Metrika*, 33(1), 321–336.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2), 383–386.
- Hommel, G. (1989). A comparison of two modified Bonferroni procedures. *Biometrika*, 76(3), 624–625.
- Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(suppl\_1), S96–S104.
- Marcus, R., Peritz, E., & Gabriel, K. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3), 655–660.
- Rødland, E. A. (2006). Simes' procedure is 'valid on average. *Biometrika*, 93(3), 742–746.
- Sarkar, S. (2008). On the Simes inequality and its generalization. *IMS Collections Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 1, 231–242.
- Schroeder, M., Haibe-Kains, B., Culhane, A., Sotiropoulos, C., Bontempi, G., & Quackenbush, J. (2011). breastCancerNKI: Genexpression dataset published by van't Veer et al. [2002] and van de Vijver et al. [2002] (NKI). R package version 1.14.0.
- Sham, P. C. & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15(5), 335–346.

- Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751–754.
- Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 1–25.
- Tamhane, A. C. & Gou, J. (2018). Advances in P-value based multiple test procedures. *Journal of Biopharmaceutical Statistics*, 28(1), 10–27.
- Van de Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., ... Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25), 1999–2009.
- Wright, S. P. (1992). Adjusted P-values for simultaneous inference. *Biometrics*, 48(4), 1005–1013.

## SUPPORTING INFORMATION

Additional Supporting Information including source code to reproduce the results may be found online in the supporting information tab for this article.

**How to cite this article:** Meijer RJ, Krebs TJP, Goeman JJ. Hommel's procedure in linear time. *Biometrical Journal*. 2018;1–10. <https://doi.org/10.1002/bimj.201700316>