Online Companion for an Asymptotic Test of Optimality Conditions in Multiresponse Simulation Optimization

Ebru Angün
Department of Industrial Engineering, Galatasaray University, Ortaköy 34357 İstanbul, Turkey, eangun@gsu.edu.tr

Jack Kleijnen
Department of Information Management / CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands, kleijnen@uvt.nl

Appendix: Derivation of the Jacobian matrices of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\lambda}$

In this appendix, we show that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\lambda}$—defined in (11) and (12) in the article—are differentiable with respect to $\nabla E\left[f_{i'}\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$ for $i' \in A'\left(\mathbf{d}_1\right)$. Apparently, $\boldsymbol{\varepsilon}$ is everywhere partially differentiable with respect to $\nabla E\left[f_0\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$:

$$\partial \boldsymbol{\varepsilon} = \left\{\mathbf{I}_k - \boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right\}\partial E\left[f_0\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right] \tag{1}$$

where the $k \times k$ matrix within the curly brackets denotes the first partial derivative of $\boldsymbol{\varepsilon}$ with respect to $\nabla E\left[f_0\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$. To show that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\lambda}$ are also partially differentiable with respect to $\nabla E\left[f_{j'}\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$, we use the following theorem in Magnus and Neudecker (1988, p. 151).

Theorem 1 *Let T be the set of non-singular real $m \times m$ matrices. Let S be an open subset of $R^{n\times q}$. If the matrix function $F : S \to T$ is k times (continuously) differentiable on S, then so is the matrix function $F^{-1} : T \to S$ defined by $F^{-1}\left(\mathbf{X}\right) = \left(F\left(\mathbf{X}\right)\right)^{-1}$, and the differential is $dF^{-1} = -F^{-1}\left(dF\right)F^{-1}$.*

Because $\nabla E\left[f_{j'}\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$ for $j' \in A\left(\mathbf{d}_1\right)$ are columns of the matrix $\boldsymbol{\Gamma}$, this $\boldsymbol{\Gamma}$ is everywhere partially differentiable with respect to $\nabla E\left[f_{j'}\left(\mathbf{d}_1, \boldsymbol{\omega}_1\right)\right]$. Furthermore, because of our assumption of linearly independent binding constraints at $\mathbf{d}_1$, $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}$ is non-singular. Using

the above theorem, the partial differential of $\boldsymbol{\varepsilon}$ with respect to $\nabla E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]$ is given by

$$
\begin{aligned}
\partial\boldsymbol{\varepsilon} &= -\partial\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]+\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\partial\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1} \\
&\quad \boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial\boldsymbol{\Gamma}\right)^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad = -\partial\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]+\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left[\left(\partial\boldsymbol{\Gamma}\right)^T\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^T\partial\boldsymbol{\Gamma}\right]\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1} \\
&\quad \boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial\boldsymbol{\Gamma}\right)^T\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad = -\partial\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]+\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial\boldsymbol{\Gamma}\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial\boldsymbol{\Gamma}\right)^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right].
\end{aligned}
\tag{2}
$$

Note that $\partial\boldsymbol{\Gamma}=\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T$, where $\mathbf{e}_{j'}$ is the $\left|A\left(\mathbf{d}_1\right)\right|\times 1$ vector of zeros, except a one in the j'th position. Replacing $\partial\boldsymbol{\Gamma}$ by $\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T$, we can rewrite (2) as follows:

$$
\begin{aligned}
\partial\boldsymbol{\varepsilon} &= -\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T\right)\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T\right)\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad -\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\mathbf{e}_{j'}^T\right)^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad = -\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\left[\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right] \\
&\quad -\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\left[\left(\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right)^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right] \\
&\quad = -\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right]\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad +\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad +\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right]\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad -\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]^T\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \\
&\quad = \left\{-\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]\right]\left(\mathbf{I}_k-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right)\right. \\
&\quad \left.-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]^T\left(\mathbf{I}_k-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right)\right\}\partial E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]
\end{aligned}
\tag{3}
$$

where the expressions within the square brackets are scalars, and the $k\times k$ matrix within the curly brackets on the last two lines is the first partial derivative of $\boldsymbol{\varepsilon}$ with respect to $\nabla E\left[f_{j'}\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]$.

Analogously, the partial differential of $\boldsymbol{\lambda}$ with respect to $\nabla E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right]$ is given by

$$
\partial\boldsymbol{\lambda}=\left\{\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right\}\partial E\left[f_0\left(\mathbf{d}_1,\,\boldsymbol{\omega}_1\right)\right] \tag{4}
$$

where the $|A(\mathbf{d}_1)| \times k$ matrix within the curly brackets denotes the first partial derivative of $\boldsymbol{\lambda}$ with respect to $\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]$. Furthermore, the partial differential of $\boldsymbol{\lambda}$ with respect to $\nabla E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]$ is given by

$$
\begin{aligned}
\partial\boldsymbol{\lambda} &= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\partial\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}(\partial\boldsymbol{\Gamma})^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left[(\partial\boldsymbol{\Gamma})^T\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^T\partial\boldsymbol{\Gamma}\right]\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}(\partial\boldsymbol{\Gamma})^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}(\partial\boldsymbol{\Gamma})^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad - \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}(\partial\boldsymbol{\Gamma})^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)].
\end{aligned} \tag{5}
$$

Again replacing $\partial\boldsymbol{\Gamma}$ by $\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\,\mathbf{e}_{j'}^T$, we rewrite (5) as follows:

$$
\begin{aligned}
\partial\boldsymbol{\lambda} &= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\,\mathbf{e}_{j'}^T\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad - \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\left(\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\,\mathbf{e}_{j'}^T\right)\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\,\mathbf{e}_{j'}^T \\
&\quad + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\left(\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\,\mathbf{e}_{j'}^T\right)^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\left[\left(\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right)^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right] \\
&\quad - \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right] \\
&\quad + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\left[\left(\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right)^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right] \\
&= -\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]^T\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad - \left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right]\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&\quad + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]^T\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)] \\
&= \left\{-\left[\mathbf{e}_{j'}^T\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\right]\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right. \\
&\quad \left. + \left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_{j'}\nabla E[f_0(\mathbf{d}_1, \boldsymbol{\omega}_1)]\left(\mathbf{I}_k - \boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^T\right)\right\}\partial E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]
\end{aligned} \tag{6}
$$

where the expressions within the square brackets are scalars, and the $|A(\mathbf{d}_1)| \times k$ matrix within the curly brackets on the last two lines is the first partial derivative of $\boldsymbol{\lambda}$ with respect to $\nabla E[f_{j'}(\mathbf{d}_1, \boldsymbol{\omega}_1)]$.

Now, $\nabla\widehat{\boldsymbol{\varepsilon}}_N$, which is a consistent estimator of the $(k\,|A'(\mathbf{d}_1)|) \times k$ Jacobian matrix of $\boldsymbol{\varepsilon}$, is formed by the transposes of the matrices within the curly brackets in (1) and in (3) for each $j' \in A(\mathbf{d}_1)$, and $\nabla\widehat{\boldsymbol{\lambda}}_N$, which is a consistent estimator of the $(k\,|A'(\mathbf{d}_1)|) \times |A(\mathbf{d}_1)|$ Jacobian matrix of $\boldsymbol{\lambda}$, is formed by the transposes of the matrices within the curly brackets in (4) and in (6) for each $j' \in A(\mathbf{d}_1)$.

Reference

Magnus, J.R., H. Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, England.

Further Numerical Results

Toy problem with an unconstrained optimum

To obtain an unconstrained minimizer, we replace the objective function of the toy problem in (30) in the article by

$$\begin{array}{ll} \text{minimize} & E\left[(d_1-2)^2+5(d_2+2.2)^2+\epsilon_0\right] \\ \text{subject to} & E\left[-(d_1-3)^2-d_2^2-d_1d_2+\epsilon_1\right]\geq -4 \\ & E\left[-d_1^2-3\left(d_2+1.061\right)^2+\epsilon_2\right]\geq -9 \end{array} \tag{7}$$

where now the optimal input is $(2,\ -2.2)^T$ with a mean objective value of 0.

We experiment with two different local areas with the central points $(2,\ -2.2)^T$ and $(2.5,\ -1.4)^T$. We use 5% as the user-defined percentage to determine the inputs in the CCD design, R and $2R$ to determine the axial points, and 250 and 2500 replicates for all cases. For the local area with the central point $(2,\ -2.2)^T$ we select $\sigma_{0,0}=25$, $\sigma_{1,1}=9$, and $\sigma_{2,2}=16$. For the local area around $(2.5,\ -1.4)^T$ we select $\sigma_{0,0}=6$, $\sigma_{1,1}=4$, and $\sigma_{2,2}=3$. For both local areas we select $\rho_{0,1}=-0.2$, $\rho_{0,2}=0.7$, and $\rho_{1,2}=-0.4$. Because there is no constraint binding at $(2,\ -2.2)^T$ and $(2.5,\ -1.4)^T$, Stage 2 of the test is skipped and the type-I error probabilities are allocated as $\alpha_1=5\%$ and $\alpha_3=5\%$.

Tables 1 and 2 summarize the results of 2000 macro-replicates of the two stages of our test for the local areas centered at $(2,\ -2.2)^T$ and $(2.5,\ -1.4)^T$ respectively. We offer the following comments.

1. In Table 1, the estimates of type-II error rates of the t-test are $33/2000=0.0165$ (rep. $=250$, axial $=R$, and axial $=2R$), and zero (rep. $=2500$, axial $=R$, and axial $=2R$); see the second column. In Table 2, they are all zero. This is to be expected because $(2.5,\ -1.4)^T$ is located further away from both constraints than $(2,\ -2.2)^T$ is.

2. In Table 1, the estimated conditional type-I error rates are all lower than the nominal $\alpha_3=0.05$; see the third column. The estimated unconditional type-I error rates are

96/2000 = 0.048, 84/2000 = 0.042, 98/2000 = 0.049, and 91/2000 = 0.0455, which are also all lower than $\alpha_3 = 0.05$

3. In Table 2, increasing the number of replicates from 250 to 2500 does improve the power of the KKT test. The estimated type-II error rates decrease from 70/2000 = 0.035 (rep. = 250, axial = R) to 2/2000 = 0.001 (rep. = 2500, axial = R), and from 61/2000 = 0.0305 (rep. = 250, axial = $2R$) to 0 (rep. = 2500, axial = $2R$).

Toy problem with a constrained optimum

To determine the number of replicates at each design point for the homogeneous and heterogeneous variance cases, we repeat the experiments using all data in Sections 4.1.1 and 4.1.2 of the article except that after each 2000 macro-replicates, we increase the number of replicates by 100—starting from 250 up to and including 2550 replicates—and rerun the program. We illustrate the resulting estimated type-I error rates of the KKT test (Stage 3) in Figures 1 and 2 for the homogeneous and heterogeneous cases, respectively; the prespecified type-I error rate for the KKT test is $\alpha_3 = 4\%$. Examining both figures where the estimated type-I error rates are all below $\alpha_3 = 4\%$, we decide to use both $m_l = 250$ (small number of replicates) and $m_l = 850$ (big number of replicates) for both the homogeneous and the heterogeneous cases.

After fixing the number of replicates to 250 and then to 850, we repeat the experiments using all data in Section 4.1.1 of the article except that the user-defined percentage to determine the size of the local experimental area is now 1.25%. We summarize our results in Tables 3 through 5. Below, we compare these results with those obtained when we use 5% as the user-defined percentage (with the results in Tables 2 through 4 in the article).

1. Comparing the second column of Tables 3 through 5 with the one in Tables 2 through 4 in the article, it is clear that the estimated type-I error rates of the t-test are not affected by the decrease in the local area size.

2. Comparing the fourth column of Table 3 with the one in Table 2 in the article, we find that Stage 3, which tests the KKT conditions, gives better results (i.e., rejects the KKT conditions less often at the true optimal point) for large local areas (size = 5%) than for small local areas (size = 1.25%). A tentative explanation can be given as follows. Larger local areas give larger OLS estimates $\widehat{\boldsymbol{\beta}}_i$ in absolute magnitude. Furthermore,

the increase in the local area size decreases the variances of the OLS estimators; see the equation (4) in the article. Then, larger signals (the numerator in equation (6) of the article) and smaller noises (the denominator in equation (6) of the article) result in larger average signal/noise ratios in Table 6 in absolute magnitude, which improves the performance of our stopping rule.

3. Comparing the third column of Table 3 with the one in Table 2 in the article clearly shows that Stage 2 (which tests the condition number of the matrix $\mathbf{\Gamma}$ of the gradients of the binding constraint) is very badly affected by the decrease in the local area size. This problem is solved by increasing the number of replicates from 250 to 850.

4. Comparing the third column of Table 4 with the one in Table 3 in the article shows that the type-II error estimates are slightly worse for small local areas (size = 1.25%) than for large local areas (size = 5%) in case of a small number of replicates (rep. = 250). The difference vanishes with the increase in the number of replicates. Furthermore, the average signal/noise ratios in Table 7 have similar magnitudes.

5. Comparing the third column of Table 5 with the one in Table 4 in the article shows that the type-II error estimates are again worse for small local areas (size = 1.25%) than for large local areas (size = 5%) in case of a small number of replicates (rep. = 250), and the difference again vanishes with the increase in the number of replicates. The average signal/noise ratios in Table 8 have again similar magnitudes.

In general, we conclude that when it is possible to obtain unbiased estimators of the gradients, large local areas give better results. The change in performance of the stopping rule in larger local areas becomes bigger as the point tested lies closer to the true optimal point.

(s, S) inventory problem with a service-level constraint

To determine the number of replicates at each design point, we repeat the experiments using all data in Section 4.2 of the article except that after each 1000 macro-replicates, we increase the number of replicates by one—starting from two up to and including ten replicates—and rerun the program. The resulting estimated type-I error rates of the KKT test are illustrated in Figures 3 and 4, where 2.5% and 1.25% are used as the user-defined percentages, respectively, and the prespecified type-I error rate is $\alpha_3 = 5\%$. In Figure 3,

the estimated type-I error rates are all above $\alpha_3 = 5\%$ with larger local experimental areas ($2R$ to determine axial points) giving worse results than smaller ones (R to determine axial points). We have similar observations for the estimates in Figure 4. However, in Figure 4, for smaller local experimental areas, the estimated type-I error rates are slightly above $\alpha_3 = 5\%$, and getting very close to $\alpha_3 = 5\%$ starting from five replicates. Because 1000 macro-replicates of the inventory problem with five replicates per design point take about eight CPU hours, we decide to fix $m_l = 2$, where now these 1000 macro-replicates take only three CPU hours.

In Table 9, we present the average signal/noise ratios and their sample standard deviations computed from 1000 macro-replicates. These signal/noise ratios have similar absolute magnitudes; increasing the experimental area size does not increase the signal/noise ratios whereas Table 6 shows that they did. Furthermore, the signal/noise ratios in Table 9 may be biased (see the discussion in Section 4.2 of the article); this bias decreases when the size of the local experimental area decreases. The least biased estimators of the signal/noise ratios are obtained for size = 1.25% and axial = R, while the most biased ones are obtained for size = 2.5% and axial = $2R$. We conclude that the bias affects our stopping rule very adversely; i.e., we obtain the best results for the smallest local area with size = 1.25% and axial = R and the worst results for the largest local area with size = 2.5% and axial = $2R$; see Section 4.2 in the article.

Table 1: Estimated type-I error rates for the local area with the true unconstrained optimal point as the central point $(2, -2.2)^T$

local area	no constraint found active	KKT rejected
size = 5%, rep. = 250 axial = R	1967/2000=0.9835	96/1967=0.0488
size = 5%, rep. = 2500 axial = R	2000/2000=1	84/2000=0.042
size = 5%, rep. = 250 axial = $2R$	1967/2000=0.9835	98/1967=0.0498
size = 5%, rep. = 2500 axial = $2R$	2000/2000=1	91/2000=0.0455

Table 2: Estimated power for the local area away from the true unconstrained optimum centered at $(2.5, -1.4)^T$

local area	no constraint found active	KKT rejected
size = 5%, rep. = 250 axial = R	2000/2000=1	1230/2000=0.615
size = 5%, rep. = 2500 axial = R	2000/2000=1	1998/2000=0.0999
size = 5%, rep. = 250 axial = $2R$	2000/2000=1	1339/2000=0.6695
size = 5%, rep. = 2500 axial = $2R$	2000/2000=1	2000/2000=1

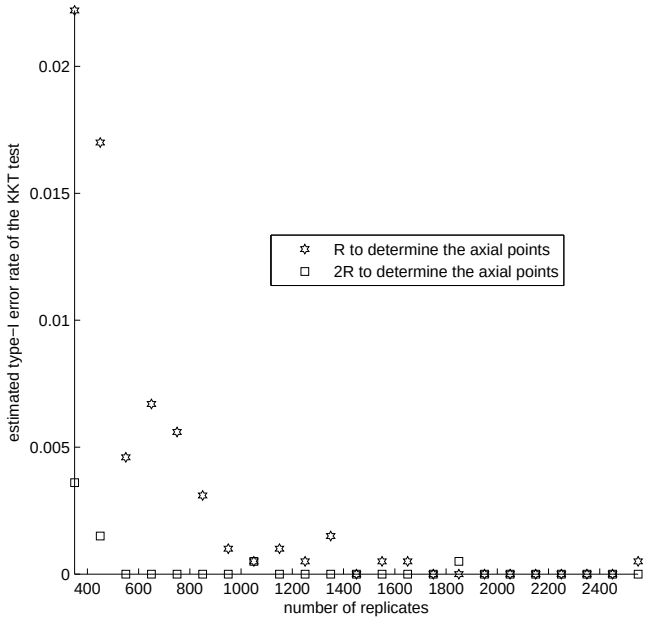

Figure 1: Estimated type-I error rates for the KKT test versus the number of replicates: homogeneous case and $\alpha_3 = 4\%$

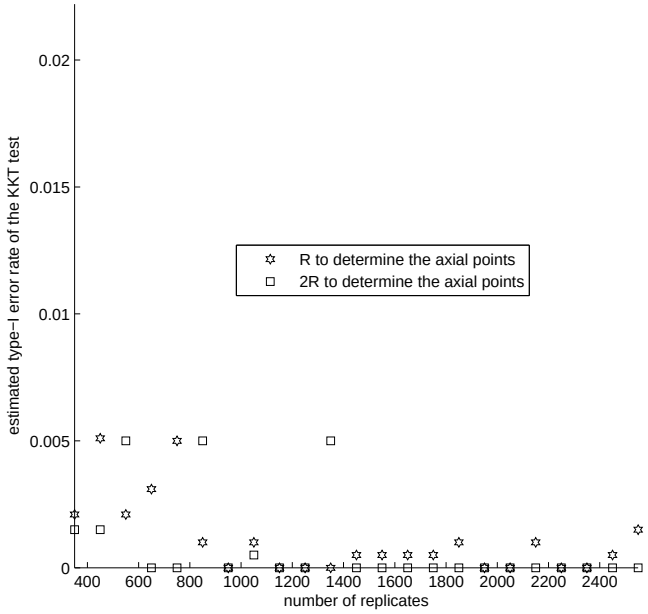

Figure 2: Estimated type-I error rates for the KKT test versus the number of replicates: heterogeneous case and $\alpha_3 = 4\%$

Table 3: Estimated type-I error rates for the local area with the true optimal point as the central point $(1.6458, -2.5091)^T$ and locally homogeneous variances

local area	both constraints active	$\mathbf{\Gamma}$ found ill-conditioned	KKT rejected
size = 1.25%, rep. = 250 axial = R	1955/2000=0.9775	584/1955=0.2987	58/1371=0.0423
size = 1.25%, rep. = 850 axial = R	1960/2000=0.9800	59/1960=0.0301	48/1901=0.0252
size = 1.25%, rep. = 250 axial = $2R$	1947/2000=0.9735	387/1947=0.1988	37/1560=0.0237
size = 1.25%, rep. = 850 axial = $2R$	1946/2000=0.9730	0/1946=0	17/1946=0.0087

Table 4: Estimated power for the local area furthest away from the true optimum centered at $(1, -1)^T$ and locally homogeneous variances

local area	constraint 1 active	KKT rejected
size = 1.25%, rep. = 250 axial = R	1953/2000=0.9765	1769/1953=0.9058
size = 1.25%, rep. = 850 axial = R	1946/2000=0.9730	1946/1946=1
size = 1.25%, rep. = 250 axial = $2R$	1953/2000=0.9765	1949/1953=0.9980
size = 1.25%, rep. = 850 axial = $2R$	1946/2000=0.9730	1946/1946=1

Table 5: Estimated power for the intermediate local area centered at $(1.8,\ -2.4466)^T$ and locally homogeneous variances

local area	constraint 2 active	KKT rejected
size = 1.25%, rep. = 250 axial = R	1962/2000=0.9810	1113/1962=0.5673
size = 1.25%, rep. = 850 axial = R	1950/2000=0.9750	1845/1950=0.9462
size = 1.25%, rep. = 250 axial = $2R$	1962/2000=0.9810	1526/1962=0.7778
size = 1.25%, rep. = 850 axial = $2R$	1950/2000=0.9750	1928/1950=0.9887

Table 6: Average signal/noise ratios over 2000 macro-replicates and their sample standard deviations for the local area centered at $(1.6458,\ -2.5091)^T$: homogeneous case

local area	$\left(\overline{\widehat{\gamma}}_{i;1;2}, \overline{\widehat{\gamma}}_{i;1;1}, \overline{\widehat{\gamma}}_{i;2;2}\right)$	st. deviations
size = 5%, rep. = 250	$i = 0$, (-0.0157, -0.0281, 0.0818)	(0.0223, 0.0222, 0.0222)
axial = R	$i = 1$, (-0.0419, -0.0388, -0.0295)	(0.0225, 0.0224, 0.0223)
	$i = 2$, (0.0029, -0.0272, -0.0511)	(0.0225, 0.0222, 0.0224)
size = 5%, rep. = 850	$i = 0$, (0.0403, 0.0756, 0.2331)	(0.0225, 0.0224, 0.0224)
axial = R	$i = 1$, (-0.0745, -0.1359, -0.1324)	(0.0219, 0.0223, 0.0225)
	$i = 2$, (0.0223, -0.0028, -0.1654)	(0.0223, 0.0224, 0.0221)
size = 1.25%, rep. = 250	$i = 0$, (-0.0114, 0.0149, 0.0171)	(0.0228, 0.0222, 0.0223)
axial = R	$i = 1$, (-0.0116, 0.0063, 0.0018)	(0.0223, 0.0217, 0.0220)
	$i = 2$, (-0.0019, -0.0016, 0.0022)	(0.0224, 0.0219, 0.0218)
size = 1.25%, rep. = 850	$i = 0$, (0.0216, -0.0007, 0.0208)	(0.0220, 0.0221, 0.0225)
axial = R	$i = 1$, (0.0140, -0.0214, 0.0013)	(0.0224, 0.0224, 0.0222)
	$i = 2$, (-0.0055, 0.0021, -0.0252)	(0.0218, 0.0221, 0.0223)

Table 7: Average signal/noise ratios over 2000 macro-replicates and their sample standard deviations for the local area centered at $(1, -1)^T$: homogeneous case

local area	$\left(\overline{\widehat{\gamma}}_{i;1;2}, \overline{\widehat{\gamma}}_{i;1;1}, \overline{\widehat{\gamma}}_{i;2;2}\right)$	st. deviations
size = 5%, rep. = 250	$i = 0$, (-0.0143, -0.0074, -0.0105)	(0.0226, 0.0222, 0.0222)
axial = R	$i = 1$, (-0.0150, -0.0090, -0.0083)	(0.0225, 0.0220, 0.0221)
size = 5%, rep. = 850	$i = 0$, (0.0339, 0.0193, 0.0729)	(0.0220, 0.0224, 0.0222)
axial = R	$i = 1$, (-0.0179, -0.0283, -0.0169)	(0.0224, 0.0219, 0.0220)
size = 1.25%, rep. = 250	$i = 0$, (-0.0143, -0.0105, -0.0263)	(0.0226, 0.0222, 0.0222)
axial = R	$i = 1$, (-0.0088, -0.0037, -0.0030)	(0.0225, 0.0220, 0.0221)
size = 1.25%, rep. = 850	$i = 0$, (0.0339, 0.0093, 0.0229)	(0.0220, 0.0224, 0.0222)
axial = R	$i = 1$, (0.0016, -0.0116, -0.0002)	(0.0224, 0.0219, 0.0220)

Table 8: Average signal/noise ratios over 2000 macro-replicates and their sample standard deviations for the local area centered at $(1.8, -2.4466)^T$: homogeneous case

local area	$\left(\overline{\widehat{\gamma}}_{i;1;2}, \overline{\widehat{\gamma}}_{i;1;1}, \overline{\widehat{\gamma}}_{i;2;2}\right)$	st. deviations
size = 5%, rep. = 250	$i = 0$, (-0.0623, 0.0375, 0.0830)	(0.0225, 0.0217, 0.0228)
axial = R	$i = 2$, (-0.0357, -0.0009, -0.0589)	(0.0226, 0.0223, 0.0226)
size = 5%, rep. = 850	$i = 0$, (-0.0135, 0.0820, 0.2487)	(0.0228, 0.0219, 0.0216)
axial = R	$i = 2$, (-0.0162, -0.0369, -0.1757)	(0.0223, 0.0216, 0.0219)
size = 1.25%, rep. = 250	$i = 0$, (-0.0623, 0.0233, 0.0079)	(0.0225, 0.0217, 0.0228)
axial = R	$i = 2$, (-0.0357, 0.0169, -0.0026)	(0.0226, 0.0223, 0.0226)
size = 1.25%, rep. = 850	$i = 0$, (-0.0135, 0.0370, 0.0115)	(0.0228, 0.0219, 0.0216)
axial = R	$i = 2$, (-0.0162, 0.0194, 0.0022)	(0.0223, 0.0216, 0.0219)

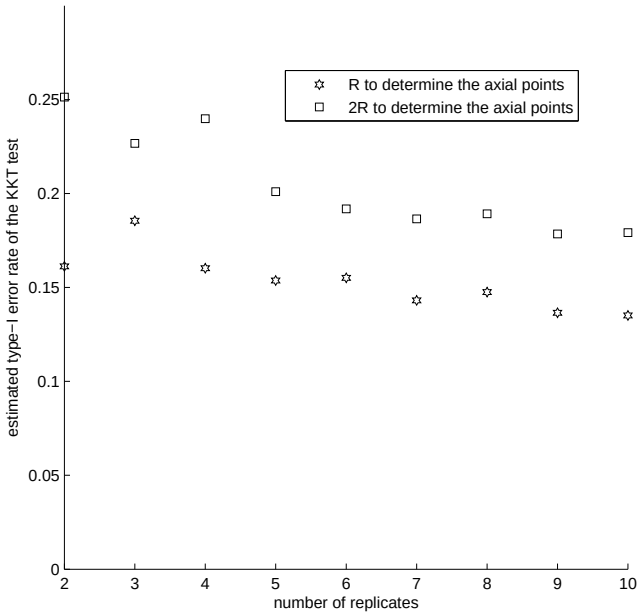

Figure 3: (s, S) inventory problem: estimated type-I error rates for the KKT test versus the number of replicates ($\alpha_3 = 5\%$ and 2.5% as the user-defined percentage)

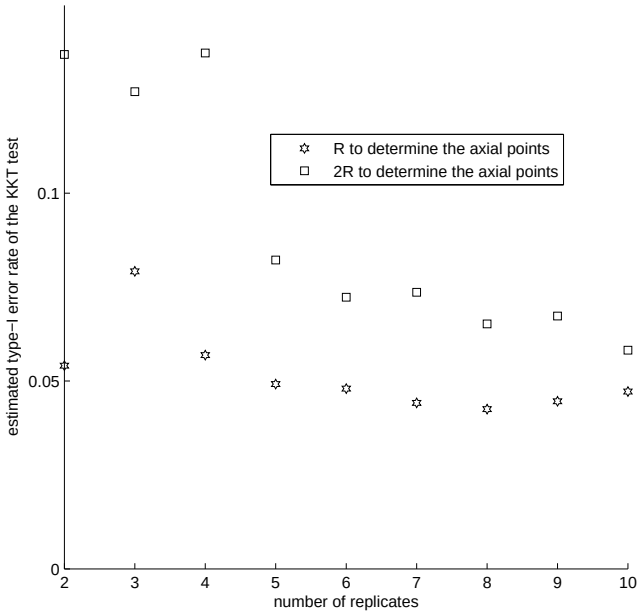

Figure 4: (s, S) inventory problem: estimated type-I error rates for the KKT test versus the number of replicates ($\alpha_3 = 5\%$ and 1.25% as the user-defined percentage)

Table 9: $(s,\ S)$ inventory problem: average signal/noise ratios over 1000 macro-replicates and their sample standard deviations for the local area centered at $(1020,\ 1075)^T$ ($i = 0$: expected cost, $i = 1$: expected fill-rate)

local area	$\left(\overline{\overline{\gamma}}_{i;1;2},\ \overline{\overline{\gamma}}_{i;1;1},\ \overline{\overline{\gamma}}_{i;2;2}\right)$	st. deviations
size = 2.5%, periods = 2500	$i = 0$, (0.0260, -0.0587, -0.0154)	(0.0341, 0.0350, 0.0342)
axial = R	$i = 1$, (0.0989, -0.0304, -0.0109)	(0.0337, 0.0334, 0.0348)
size = 2.5%, periods = 2500	$i = 0$, (-0.0166, 0.0312, 0.0280)	(0.0350, 0.0353, 0.0353)
axial = $2R$	$i = 1$, (0.0314, -0.2059, 0.0005)	(0.0334, 0.0367, 0.0365)
size = 1.25%, periods = 2500	$i = 0$, (-0.0536, 0.0106, -0.0086)	(0.0343, 0.0346, 0.0336)
axial = R	$i = 1$, (-0.0301, -0.0189, 0.0066)	(0.0342, 0.0347, 0.0344)
size = 1.25%, periods = 2500	$i = 0$, (0.0090, -0.0128, -0.0584)	(0.0352, 0.0336, 0.0342)
axial = $2R$	$i = 1$, (0.0011, -0.1152, -0.0147)	(0.0338, 0.0341, 0.0332)