AN EXAMPLE ARTICLE*

DIANNE DOE†, PAUL T. FRANK‡, AND JANE E. SMITH‡

Abstract. This is an example SIAM LaTeX article. This can be used as a template for new articles. Abstracts must be able to stand alone and so cannot contain citations to the paper's references, equations, etc. An abstract must consist of a single paragraph and be concise. Because of online formatting, abstracts must appear as plain as possible. Any equations should be inline.

Key words. example, LaTeX

AMS subject classifications. 68Q25, 68R10, 68U05

1. Introduction.

1.1. Moment Matching Estimation.

Consider an N dimensional independent sample $X^{(1)},...,X^{(T)}$ generated by a stochastic process that is governed by a joint probability density function (pdf) $f_X(x) : \mathbb{R}_+^N \to \mathbb{R}_+$. The pdf is parameterized by a P vector θ whose domain Θ is a subset of the Euclidean space. The problem is to identify the model parameter θ of density function $f_X(x;\theta)$ using an M-dimensional independent sample $Y^{(1)},...,Y^{(T)}$, whose relation with sample of X is given by a known function $Y^{(t)} = g(X^{(t)}), g : \mathbb{R}_+^N \to \mathbb{R}_+^M$ for $t = 1,..,T$.

An initial attempt might be to first solve for $X^{(t)} = g^{-1}(Y^{(t)})$ and then to find the estimate of θ based on $X^{(1)},...,X^{(T)}$. Unfortunately, function g is often not injective due to O-D flow observability issue; multiple solutions of $X^{(t)}$ may be admitted by solving $g(X^{(t)}) - Y^{(t)} = 0$ in some or all t. Simultaneous determination of all $X^{(1)},...,X^{(T)}$ from $Y^{(1)},...,Y^{(T)}$ may alleviate the observatbility issues as reported in [?] and [?]. However, this additional layer of treatment could significantly complicate the problem to be analyzed. One may also hope to be able to specify $f_Y(y|\theta)$, which bestows the "full" amount of information on observed data and serves as the premise of ML estimation and related hypothesis tests. Following the fact that the probability contained in a differential area must be invariant under the change of variables, the pdf of Y is the sum of $\left|\frac{d}{dy}(g^{-1}(y))\right| \cdot f_X(g^{-1}(y))$ for all points that solve the inverse problem. However, calculating the derivative of $g^{-1}(y)$ is not a straightforward task.

Fortunately, based on the law of the unconscious statistician, the calculation of the moments of Y conditional to θ does not involve anything related to $g^{-1}(y)$:

$$E(h(Y)|\theta) = \int_{\mathbb{R}_+^N} h(g(x)) f_X(x|\theta) dx. \tag{1.1}$$

where, for now, $h(\cdot) : \mathbb{R}_+^M \to \mathbb{R}^R$ can be considered as a function defined on entire Y space for calculating various orders of moments. Its specific form will be given later. According to the analogy principle, an estimator of θ can be obtained by solving for θ that makes the sample analogs of population moments small. Therefore, the density estimation problem becomes finding θ using the sample moments of Y calculated from $Y^{(1)},...,Y^{(T)}$. The details of our method is presented in the following section.

---

*Submitted to the editors DATE.

Funding: This work was funded by the Fog Research Institute under contract no. FRI-454.

†Imagination Corp., Chicago, IL (ddoe@imag.com, http://www.imag.com/~ddoe/).

‡Department of Applied Mathematics, Fictional University, Boise, ID (ptfrank@fictional.edu, jesmith@fictional.edu).

Let us first introduce notation that is convenient for defining moments for a multivariate random variable. Define $r \in \mathbb{Z}_+^M$ as an order indicator vector for a particular moment $\mu_{\{r\}}$ of multivariate random variable Y. The j-th element r_j in vector r specifies the order of Y_j. For instance, when $M = 5$, $r = [1,0,0,0,0]^{\mathrm{T}}$ indicates the mean of Y_1 and $r = [0,1,1,0,0]^{\mathrm{T}}$ indicates the second moment of Y_2 and Y_3.

Now $h(\cdot)$ in (1.1) can be defined as a $R \times 1$ vector function with its entry

$$h_{\{r\}}(y) = \prod_{j=1}^{M} y_j^{r_j}. \tag{1.2}$$

The moment vector functions $\mu(\theta)$ is the expectation of $h(y)$ conditioning on θ, that is

$$\mu(\theta) = E\left(h(Y)\middle|\theta\right) = \int_{\mathbb{R}_+^N} h(g(x)) f_X(x|\theta) dx. \tag{1.3}$$

Note that the order of moment $\mu_{\{r\}}$ is $\sum_{j=1}^{M} r_j$. The vector μ up to $\bar{r}$-th order consists of all $\mu_{\{r\}}$'s such that $0 < \sum_{j=1}^{M} r_j \leq \bar{r}$ and is of length

$$R = \sum_{a=1}^{\bar{r}} \binom{M+a-1}{a}, \tag{1.4}$$

where $\binom{n}{k}$ is the number of k-combinations from a given set of n elements.

Now we define functions γ as

$$\gamma(Y,\theta) = \mu(\theta) - h(Y) = \int_{\mathbb{R}_+^M} f_X(x|\theta) h(g(x)) dx - h(Y). \tag{1.5}$$

Let θ_0 be the true value of θ in the population. A generalized method of moment (GMM) requires $\gamma(Y,\theta)$ to have a population expected value equals to a zero vector, if $\theta = \theta_0$:

$$E(\gamma(Y,\theta)) = 0, \text{ if } \theta = \theta_0 \tag{1.6}$$

When the number of estimated parameters is less than the number of equations in (1.6), i.e. $P < R$, in such an over-identification case, the sample mean of $\gamma(Y,\theta)$, denoted as $\gamma_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \gamma(Y^{(t)},\theta)$, may not have a root θ for all elements in $\gamma_T(\theta)$ simultaneously equal to zero. Then a GMM estimator of θ_0, θ_T, is a vector function of data that minimizes the weighted quadratic distance of sample generalized moments $\gamma_T(Y,\theta)$ from zeros, that is

(1.7)
$$\text{(Exact Density Estimation)} \quad J_T(\theta) \equiv T\gamma_T(\theta)^{\mathrm{T}} W \gamma_T(\theta), \text{ with } \theta_T = \arg\min_{\theta\in\Theta} J_T(\theta).$$

where W is a $R \times R$ symmetric weighting matrix. For brevity, we use the subscript T to claim the dependence of a vector on a sample $Y^{(1)}, ..., Y^{(T)}$. With sample moments $h_T = \frac{1}{T}\sum_{t=1}^{T} h(Y^{(t)})$, the problem (1.7) is equivalent to the following generalized ℓ_2 norm minimization problem,

$$J_T(\theta) = T(\mu(\theta) - h_T)^{\mathrm{T}} W(\mu(\theta) - h_T). \tag{1.8}$$

Note with $\bar{r} = 1$, the GMM method is the same with GLS approach used in several travel demand estimation studies (e.g., [**?**]).

One critical step of finding a GMM estimator is the determination of "distance metric" W_T. Define a dispersion matrix

$$\begin{aligned}\Omega(\theta) &\equiv E(\gamma(\theta, Y)\gamma(\theta, Y)^{\mathrm{T}}) \\ &= \int_{\mathrm{IR}_+^N} h(g(x))h(g(x))^{\mathrm{T}} f_X(x|\theta)dx - \int_{\mathrm{IR}_+^N} h(g(x))f_X(x|\theta)dx \int_{\mathrm{IR}_+^N} h(g(x))^{\mathrm{T}} f_X(x|\theta)dx\end{aligned} \tag{1.9}$$

Then the best choice for W is $W(\theta_0) = (\Omega(\theta_0))^{-1}$, which inevitably requires knowing θ_0 beforehand. In general, a suboptimal matrix $W(\tau_T)$ that depends on some sequence of "preliminary estimates" τ_T is used in place of $W(\theta_0)$. In literature of travel demand estimation, for GLS approach, [**?**] tested several alternatives of specifying W_T and compared them numerically. In general, several popular approaches to deal with this issue include two-step substitution, iterative updating and simultaneous estimation. In this specific context, because data Y and θ are separated in different terms of $\gamma(Y, \theta)$, a good candidate for W_T without involving guesstimate τ_T is actually simple to find

$$W_T = (\Omega_T)^{-1}, \text{ where } \Omega_T = \frac{1}{T}\sum_{t=1}^{T} h(Y^{(t)})h(Y^{(t)})^{\mathrm{T}} - \frac{1}{T^2}\sum_{t=1}^{T} h(Y^{(t)}) \sum_{t=1}^{T} h(Y^{(t)})^{\mathrm{T}}. \tag{1.10}$$

In order to ascertain that W_T is invertible, the linear dependency among the moments condition has to be eliminated prior to estimation. This linear dependency could stem from the relation between flows $Y_j, j = 1, ..., M$. Since the nodal balance law of traffic flows is assumed to be strictly observed at all intermediate nodes in many assignment model $g(x)$, flow observations could be linearly dependent to each other. If a flow variable Y_j is equal to a linear combination of other $Y_{j'}, \forall j' \neq j$ variables, the inclusion of any moment related to Y_j does not add any additional information for θ. In the scope of this current study, it is safe to remove such variables to obtain all linearly independent Y's without worsening estimation performance. The number of variables to remove should be the difference between M and the rank of data matrix $[Y^{(1)}, ..., Y^{(T)}]$. In case where data redundancy is preferred to preserve more information, one may employ some dimensional reduction techniques such as Principal Component Analysis, which is not covered in the current effort.

1.2. Numerical Approximation. In general, with a complicated assignment model $g(x)$ and/or an intractable probability density $f_X(x|\theta)$, the integral in the definition of $\gamma(Y, \theta)$ in (1.5) is approximated through numerical integration techniques, as an analytical evaluation may not be possible. In order to obtain desired results with the least effort, it is helpful to first construct a sufficiently large bounded region S for the majority of possible values of X to fall within. Choice of S can be based on one's preliminary knowledge about the scale of the demand.

Let $N \times 1$ real-valued vectors $Z^{[k]}, k = 1, ..., K$ be the integration points sampled from the region S and $w(z)$ is a predetermined weight function for any integration point z. Then the definite integral is approximated by the weighted sum of evaluated integrands at all integration points, that is

$$\mu(\theta) \approx \int_S h(g(x)) f_X(x|\theta) dx \approx \mu^K(\theta) \equiv \frac{1}{K} \sum_{k=1}^{K} w(Z^{[k]}) h(g(Z^{[k]})) f_X(Z^{[k]}|\theta). \tag{1.11}$$

It is important to pay attention that rather than evaluating the integral to a number, we actually employ such techniques to obtain an approximated version $\mu^K(\theta)$ of $\mu(\theta)$. Also notice in the approximation formula, $h(g(Z^{[k]}))$ is evaluated exogenously prior to the estimation, which can be conducted using parallel computing if needed. Denote $J^K(\theta)$ as the new objective function of optimization problem (1.7) after substituting the functions $\mu^K(\theta)$ into it and θ^K is the estimate based on the approximated problem, that expands to,

(1.12)

$$\text{(Approximated Density Estimation)} \quad J_T^K(\theta) \equiv T\left(\mu^K(\theta) - h_T\right)^{\mathrm{T}} W \left(\mu^K(\theta) - h_T\right)$$
$$\text{with } \theta_T^K = \arg\min_{\theta \in \Theta} J_T^K(\theta).$$

For brevity, we use the superscript K to claim the dependence of a quantity on a simulation sample $Z^{[1]}, ..., Z^{[K]}$.

Given a data sample of $Y^{(1)}, ..., Y^{(T)}$, as the number of points that the integrand functions $h(g(x)) f_X(x|\theta)$ are evaluated at rises, the estimates obtained based on the approximated GMM optimization problem will be approaching the values of those where the integrals are exactly evaluated.

Proposition 1.1 (Convergence of Approximated Density Estimator). *Suppose $f(x|\theta)$ is lower semicontinuous on $S \times \Theta$. If θ_T^K minimizes $J_T^K(\theta)$, then $\theta_T^K \to \theta_T$ possible along a sequence, then θ_T minimizes $J(\theta)$ almost surely.*

Proof. Since $h(g(x)) \geq 0, \forall x \in S$ and $f(x|\theta) \geq 0, \forall x \in S$, their product is always greater or equal to zero. Because $g(x)$ is a continuous function, the integrand $h(g(x)) f(x|\theta)$ is also lower semicontinuous on $S \times \Theta$. Then the proof is completed using uniform law of large number in functions. ∎

Most density functions are lower-semicontinuous with respect to x and θ. For those probabilistic models that have jumps in their densities, for example, uniform distribution, we simply need to assign the density to the lower branch so that lower-semicontinuity is satisfied.

As for choosing among all possible approximation techniques, we generally prefer Monte Carlo based over grid based algorithms. The reason is that the approximation error of Monte Carlo methods is not dependent on the number of X variables, while that of deterministic approaches grows exponentially on the dimension, albeit they can be much more efficient when N is small.

In the naive Monte Carlo approach, $Z^{[1]}, ..., Z^{[K]}$ is an i.i.d. sample drawn from a continuous uniform distribution defined on S. Let $V \equiv \int_S dx$ be the volume of S, then the integration weight $w(Z) = V, \forall z \in S$. By law of large numbers, $\lim_{K\to\infty} \mu^K(\theta) = \mu(\theta)$. Since $\mu^K(\theta)$ is simply the sum of i.i.d. random vectors $w(Z^{[k]}) f_X(Z^{[k]}|\theta) h(Z^{[k]})$, Lindeberg-Lévy central limit theorem ensures that

$$\sqrt{K}(\mu^K(\theta) - \mu(\theta)) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta)) \tag{1.13}$$

where

$$\Sigma(\theta) = \int_S \varphi(z) w(z)^2 f_X(z|\theta)^2 h(g(z)) h(g(z))^{\mathrm{T}} dz - \mu(\theta)\mu(\theta)^{\mathrm{T}}. \tag{1.14}$$

It can be estimated using unbiased sample covariance matrix, that is

$$\Sigma^K \equiv \frac{1}{K-1}\sum_{k=1}^{K}(w^{[k]}h(z^{[k]})f_X(z^{[k]}|\theta) - \mu^K(\theta))(w^{[k]}h(z^{[k]})f_X(z^{[k]}|\theta) - \mu^K(\theta)). \tag{1.15}$$

Often times, $f_X(x|\theta)h(x)$ has very small magnitudes on a dominant fraction of S. In this way, the computation effort of evaluating $g(x)$ on a large number of integration points is wasted then. The efficiency could be worse if S is chosen too conservative and larger than necessary. Importance sampling has a known advantage on this matter when some preliminary density estimate $\varphi_X(x)$ of X is known. In this case, the random sample $Z^1, ..., Z^{[k]}$ can be drawn from $\varphi_X(x)$ and $w^{[k]} = 1/\varphi_X(Z^{[k]})$. The naive Monte Carlo approach is a special case of importance sampling with $\varphi_X(x) = 1/V$, with the density of a uniform distribution. Importance sampling based $\mu^K(\theta)$ still has the same asymptotic distribution provided in (1.13).

2. Model Formulation.

$$\begin{aligned}
&\min_{f\in\mathbb{R}_+^K} T && \left(\mu^K(\theta) - h_T\right)^{\mathrm{T}} W \left(\mu^K(\theta) - h_T\right)\\
&\text{with} && h_T = \frac{1}{T}\sum_{t=1}^{T} h(Y^{(t)})(1.2)\\
&\text{and} && \mu^K(\theta) \equiv \frac{1}{K}\sum_{k=1}^{K} w(Z^{[k]})h(g(Z^{[k]}))f^{[k]}\\
&\text{subject to} && |f^{[k]} - f^{[l]}| \le \tau(||Z^{[k]} - Z^{[l]}||_2), \quad k,l = 1,...,K, Z^{[l]} \in \mathcal{N}(Z^{[k]})
\end{aligned} \tag{2.1}$$

where τ is a controlling parameter that needs to be carefully chosen. The distance does not have to be Euclidean distance. $\mathcal{N}$ stands for a neighborhood that can be specified based on user preference. We can do a fixed region or some area that covers a certain number of nearest neighbors. Let us say that the closest ten sample points around $Z^{[k]}$ are considered, so the number of constraints are ten times of that of variables.

3. Main results. We interleave text filler with some example theorems and theorem-like items.

Here we state our main result as Theorem 3.1; the proof is deferred to section SM2.

THEOREM 3.1 (LDL^T Factorization [1]). *If $A \in \mathbb{R}^{n\times n}$ is symmetric and the principal submatrix $A(1:k, 1:k)$ is nonsingular for $k = 1:n-1$, then there exists a unit lower triangular matrix L and a diagonal matrix*

$$D = \mathrm{diag}(d_1, \dots, d_n)$$

such that $A = LDL^T$. The factorization is unique.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Theorem 3.2 (Mean Value Theorem). *Suppose f is a function that is continuous on the closed interval $[a,b]$. and differentiable on the open interval (a,b). Then there exists a number c such that $a < c < b$ and*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

In other words,

$$f(b) - f(a) = f'(c)(b - a).$$

Observe that Theorems 3.1 and 3.2 and Corollary 3.3 correctly mix references to multiple labels.

Corollary 3.3. *Let $f(x)$ be continuous and differentiable everywhere. If $f(x)$ has at least two roots, then $f'(x)$ must have at least one root.*

Proof. Let a and b be two distinct roots of f. By Theorem 3.2, there exists a number c such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} = \frac{0 - 0}{b - a} = 0.$$ ∎

Note that it may require two LaTeX compilations for the proof marks to show.

Display matrices can be rendered using environments from `amsmath`:

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.1}$$

Equation (3.1) shows some example matrices.

We calculate the Fréchet derivative of F as follows:

$$F'(U,V)(H,K) = \langle R(U,V), H\Sigma V^T + U\Sigma K^T - P(H\Sigma V^T + U\Sigma K^T)\rangle \tag{3.2a}$$
$$= \langle R(U,V), H\Sigma V^T + U\Sigma K^T\rangle$$
$$= \langle R(U,V)V\Sigma^T, H\rangle + \langle \Sigma^T U^T R(U,V), K^T\rangle. \tag{3.2b}$$

Equation (3.2a) is the first line, and (3.2b) is the last line.

4. Algorithm. Sed gravida lectus ut purus. Morbi laoreet magna. Pellentesque eu wisi. Proin turpis. Integer sollicitudin augue nec dui. Fusce lectus. Vivamus faucibus nulla nec lacus. Integer diam. Pellentesque sodales, enim feugiat cursus volutpat, sem mauris dignissim mauris, quis consequat sem est fermentum ligula. Nullam justo lectus, condimentum sit amet, posuere a, fringilla mollis, felis. Morbi nulla nibh, pellentesque at, nonummy eu, sollicitudin nec, ipsum. Cras neque. Nunc augue. Nullam vitae quam id quam pulvinar blandit. Nunc sit amet orci. Aliquam erat elit, pharetra nec, aliquet a, gravida in, mi. Quisque urna enim, viverra quis, suscipit quis, tincidunt ut, sapien. Cras placerat consequat sem. Curabitur ac diam. Curabitur diam tortor, mollis et, viverra ac, tempus vel, metus.

Our analysis leads to the algorithm in Algorithm 4.1.

Curabitur ac lorem. Vivamus non justo in dui mattis posuere. Etiam accumsan ligula id pede. Maecenas tincidunt diam nec velit. Praesent convallis sapien ac est. Aliquam ullamcorper euismod nulla. Integer mollis enim vel tortor. Nulla sodales placerat nunc. Sed tempus rutrum wisi. Duis accumsan gravida purus. Nunc nunc. Etiam facilisis dui eu sem. Vestibulum semper. Praesent eu eros. Vestibulum tellus nisl, dapibus id, vestibulum sit amet, placerat ac, mauris. Maecenas et elit ut erat placerat dictum. Nam feugiat, turpis et sodales volutpat, wisi quam rhoncus neque, vitae aliquam ipsum sapien vel enim. Maecenas suscipit cursus mi.

Algorithm 4.1 Build tree

Define $P := T := \{\{1\}, \ldots, \{d\}\}$
while $\#P > 1$ **do**
  Choose $C' \in \mathcal{C}_p(P)$ with $C' := \operatorname{argmin}_{C \in \mathcal{C}_p(P)} \varrho(C)$
  Find an optimal partition tree $T_{C'}$
  Update $P := (P \backslash C') \cup \{\bigcup_{t \in C'} t\}$
  Update $T := T \cup \{\bigcup_{t \in \tau} t : \tau \in T_{C'} \backslash \mathcal{L}(T_{C'})\}$
end while
return T

5. Experimental results.

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

Figure 1 shows some example results. Additional results are available in the supplement in Table SM1.

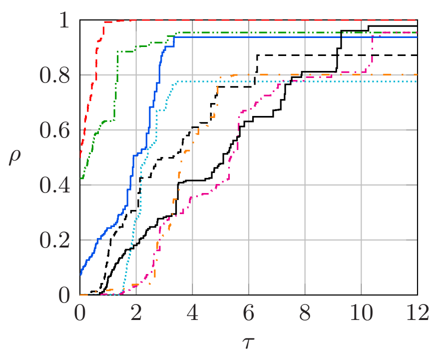

Fig. 1. *Example figure using external image files.*

Maecenas dui. Aliquam volutpat auctor lorem. Cras placerat est vitae lectus. Curabitur massa lectus, rutrum euismod, dignissim ut, dapibus a, odio. Ut eros erat, vulputate ut, interdum non, porta eu, erat. Cras fermentum, felis in porta congue, velit leo facilisis odio, vitae consectetuer lorem quam vitae orci. Sed ultrices, pede eu placerat auctor, ante ligula rutrum tellus, vel posuere nibh lacus nec nibh. Maecenas laoreet dolor at enim. Donec molestie dolor nec metus. Vestibulum libero. Sed quis erat. Sed tristique. Duis pede leo, fermentum quis, consectetuer eget, vulputate sit amet, erat.

6. Discussion of $Z = X \cup Y$.

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

7. Conclusions. Some conclusions here.

Appendix A. An example appendix. Aenean tincidunt laoreet dui. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Integer ipsum lectus, fermentum ac, malesuada in, eleifend ut, lorem. Vivamus ipsum turpis, elementum vel, hendrerit ut, semper at, metus. Vivamus sapien tortor, eleifend id, dapibus in, egestas et, pede. Pellentesque faucibus. Praesent lorem neque, dignissim in, facilisis nec, hendrerit vel, odio. Nam at diam ac neque aliquet viverra. Morbi dapibus ligula sagittis magna. In lobortis. Donec aliquet ultricies libero. Nunc dictum vulputate purus. Morbi varius. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In tempor. Phasellus commodo porttitor magna. Curabitur vehicula odio vel dolor.

LEMMA A.1. *Test Lemma.*

Acknowledgments. We would like to acknowledge the assistance of volunteers in putting together this example manuscript and supplement.

REFERENCES

[1] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 4th ed., 2013.