SUPPLEMENTARY MATERIAL TO "ESTIMATING THE RATE CONSTANT FROM BIOSENSOR DATA VIA AN ADAPTIVE VARIATIONAL BAYESIAN APPROACH"

BY YE ZHANG, ZHIGANG YAO, PATRIK FORSSÉN AND TORGNY FORNSTEDT

Chemnitz University of Technology, National University of Singapore and Karlstad University

We provide additional material of the proof of Theorem 1, finite element approximation of integral equations, as well as a demonstration of our main algorithm.

1. Proof of Theorem 1. In this material, we aim to develop the explicit formulas for carrying out each step of scheme (12) (or Algorithm 1). We first derive the necessary optimality system with respect to each component. To enforce the normalization condition of the densities $q(\mathbf{c}), q(\sigma_1^2), \cdots, q(\sigma_{N_C}^2)$, and $q(\sigma_{\mathbf{c}}^2)$, we introduce the Lagrange function $\mathfrak{L}$ for the evidence lower bound ELBO as

$$\begin{aligned}\mathfrak{L}\left(q(\mathbf{c}), q(\sigma_1^2), \cdots, q(\sigma_{N_C}^2), q(\sigma_{\mathbf{c}}^2); \lambda\right) &= \mathrm{ELBO} + \lambda_{\mathbf{c}}\left(\int q(\mathbf{c})d\mathbf{c} - 1\right) \\ &+ \sum_{j=1}^{N_C}\lambda_j\left(\int q(\sigma_j^2)d\sigma_j^2 - 1\right) + \lambda_{\sigma_{\mathbf{c}}}\left(\int q(\sigma_{\mathbf{c}}^2)d\sigma_{\mathbf{c}}^2 - 1\right),\end{aligned} \tag{1}$$

where $\lambda = \{\lambda_{\mathbf{c}}, \lambda_{\sigma_{\mathbf{c}}}\} \cup \{\lambda_j\}_{j=1}^{N_C}$ denotes the collection of Lagrange multipliers, and ELBO is defined by

$$\mathrm{ELBO} := \int q(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2) \ln\left(\frac{q(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)}{p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})}\right) d\mathbf{c}d\sigma_1^2 \cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2.$$

Taking the derivative of the Lagrange function $\mathfrak{L}$ with respect to $q(\mathbf{c})$ we obtain

$$\begin{aligned}\frac{\partial \mathfrak{L}}{\partial q(\mathbf{c})} &= \frac{\partial \mathrm{ELBO}}{\partial q(\mathbf{c})} + \lambda_{\mathbf{c}} \\ &= \int \left\{1 + \ln q(\mathbf{c}) + \sum_{j=1}^{N_C} \ln q(\sigma_j^2) + \ln q(\sigma_{\mathbf{c}}^2) - \ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})\right\} \\ &\qquad \cdot \prod_{j=1}^{N_C} q(\sigma_j^2) \cdot q(\sigma_{\mathbf{c}}^2)\ d\sigma_1^2 \cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2 + \lambda_{\mathbf{c}}\end{aligned}$$

Equating $\frac{\partial \mathfrak{L}}{\partial q(\mathbf{c})}$ to zero, and rearranging the terms in the equation, we can deduce that

$$\ln q(\mathbf{c}) = -\int \left\{ 1 + \sum_{j=1}^{N_C} \ln q(\sigma_j^2) + \ln q(\sigma_{\mathbf{c}}^2) - \ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \right\} \cdot \prod_{j=1}^{N_C} q(\sigma_j^2) \cdot q(\sigma_{\mathbf{c}}^2)\ d\sigma_1^2 \cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2 + \lambda_{\mathbf{c}}$$

Note that $\int q(\mathbf{c}) d\mathbf{c} = 1$, and all terms other than $p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$ are independent of the variable $\mathbf{c}$ and thus contribute only to the normalization condition. Hence, we deduce that at Step 3 at the k-th iteration of Algorithm 1, the component $q^k(\mathbf{c})$ can be expressed as

$$\ln q^k(\mathbf{c}) = \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \right] - \ln Z_{q^k(\mathbf{c})},$$

where the constant term $\ln Z_{q^k(\mathbf{c})}$ is given by

$$\ln Z_{q^k(\mathbf{c})} = \lambda_{\mathbf{c}} + 1 + \int \left\{ \sum_{j=1}^{N_C} \ln q^{k-1}(\sigma_j^2) + \ln q^{k-1}(\sigma_{\mathbf{c}}^2) \right\} d\sigma_1^2 \cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2.$$

By substituting the formula for $p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$, see (Zhang et al., formula (8)), into expression $\mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \right]$, we obtain

$$\begin{aligned}
&\mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \right] \\
&= \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \Big[-\frac{1}{2}(\mathbf{R} - \mathbf{Kc})^T \Sigma^{-1} (\mathbf{R} - \mathbf{Kc}) \\
&\qquad - \frac{1}{2\sigma_{\mathbf{c}}^2} (\mathbf{Lc})^T (\mathbf{Lc}) + \zeta(\sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \Big] \\
&= -\frac{1}{2}(\mathbf{R} - \mathbf{Kc})^T \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\Sigma^{-1} \right] (\mathbf{R} - \mathbf{Kc}) \\
&\qquad - \frac{1}{2} \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\sigma_{\mathbf{c}}^{-2} \right] (\mathbf{Lc})^T (\mathbf{Lc}) \\
&\qquad + \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\zeta(\sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \right],
\end{aligned}$$

where $\zeta(\sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$ contains terms in $\ln p(\mathbf{c}, \sigma_1^2, \sigma_2^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$ not involving $\mathbf{c}$. Denote by

$$\Sigma_k^{-1} = \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\Sigma^{-1} \right],\ \sigma_{\mathbf{c},k}^{-2} = \mathbb{E}_{q^{k-1}(\sigma_1^2)\cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)} \left[\sigma_{\mathbf{c}}^{-2} \right].$$

Note that for $\mathbf{c}_k = \left(\mathbf{K}^T\Sigma_k^{-1}\mathbf{K} + \sigma_{\mathbf{c},k}^{-2}\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{K}^T\Sigma_k^{-1}\mathbf{R}$, it holds

$$\begin{aligned}
-\frac{1}{2}(\mathbf{R}-\mathbf{K}\mathbf{c})^T\Sigma_k^{-1}(\mathbf{R}-\mathbf{K}\mathbf{c}) - \frac{1}{2}\sigma_{\mathbf{c},k}^{-2}(\mathbf{L}\mathbf{c})^T(\mathbf{L}\mathbf{c}) \\
&= -\frac{1}{2}(\mathbf{c}-\mathbf{c}_k)^T\left(\mathbf{K}^T\Sigma_k^{-1}\mathbf{K} + \sigma_{\mathbf{c},k}^{-2}\mathbf{L}^T\mathbf{L}\right)(\mathbf{c}-\mathbf{c}_k) \\
&\quad -\frac{1}{2}(\mathbf{R}-\mathbf{K}\mathbf{c}_k)^T\Sigma_k^{-1}(\mathbf{R}-\mathbf{K}\mathbf{c}_k) - \frac{1}{2}\sigma_{\mathbf{c},k}^{-2}(\mathbf{L}\mathbf{c}_k)^T(\mathbf{L}\mathbf{c}_k),
\end{aligned}$$

which implies that $q^k(\mathbf{c})$ follows the truncated normal distribution , i.e.

$$q^k(\mathbf{c}) = \mathcal{N}_+(\mathbf{c}_k, \left(\mathbf{K}^T\Sigma_*^{-1}\mathbf{K} + \sigma_{\mathbf{c},*}^{-2}\mathbf{L}^T\mathbf{L}\right)^{-1}).$$

Now, we derive the remaining part of the optimality system. Analogously, taking the variational derivative of the Lagrange function $\mathfrak{L}$ with respect to $q(\sigma_j^2)$ (the index j is fixed) yields

$$\begin{aligned}
&\frac{\partial\mathfrak{L}\left(q(\mathbf{c}), q(\sigma_1^2), \cdots, q(\sigma_{N_C}^2), q(\sigma_{\mathbf{c}}^2)\right)}{\partial q(\sigma_j^2)} = \frac{\partial \text{ELBO}}{\partial q(\sigma_j^2)} + \lambda_j \\
&= \int \left\{1 + \ln q(\mathbf{c}) + \sum_{j=1}^{N_C}\ln q(\sigma_j^2) + \ln q(\sigma_{\mathbf{c}}^2) - \ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})\right\} \\
&\qquad \cdot d\mathbf{c} \cdot \prod_{i\neq j} q(\sigma_i^2)\cdot q(\sigma_{\mathbf{c}}^2)\ d\mathbf{c}d\sigma_1^2\cdots d\sigma_{j-1}^2 d\sigma_{j+1}^2\cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2 + \lambda_{\mathbf{c}}
\end{aligned}$$

By the optimality condition $\frac{\partial\mathfrak{L}}{\partial q(\sigma_j^2)} = 0$ and the normalization condition $\int q(\sigma_j^2)d\sigma_j^2 = 1$, we can deduce that at the k-th iteration

$$\begin{aligned}
&\ln q^k(\sigma_j^2) = -\ln Z_{q^k(\sigma_j^2)} + \\
&\mathbb{E}_{q^k(\mathbf{c})q^k(\sigma_1^2)\cdots q^k(\sigma_{j-1}^2)q^{k-1}(\sigma_{j+1}^2)\cdots q^{k-1}(\sigma_{N_C}^2)q^{k-1}(\sigma_{\mathbf{c}}^2)}\left[\ln p(\mathbf{c}, \sigma_1^2, ..., \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})\right],
\end{aligned}$$

where the constant term $\ln Z_{q^k(\sigma_j^2)}$ is given by

$$\begin{aligned}
\ln Z_{q^k(\sigma_j^2)} = \lambda_j + 1 + \int\{\ln q^k(\mathbf{c})\left(\sum_{i<j}\ln q^k(\sigma_i^2) + \sum_{i>j}\ln q^{k-1}(\sigma_i^2)\right) \\
+ \ln q^{k-1}(\sigma_{\mathbf{c}}^2)\}d\mathbf{c}d\sigma_1^2\cdots q^{k-1}(\sigma_{j-1}^2)q^{k-1}(\sigma_{j+1}^2)\cdots d\sigma_{N_C}^2 d\sigma_{\mathbf{c}}^2.
\end{aligned}$$

Let the constant α'_j be defined by $\alpha'_j = \alpha_j + N_T/2$, and we have

$$\begin{aligned}
&\mathbb{E}_{q^k(\mathbf{c})q^k(\sigma_1^2)\cdots q^k(\sigma_{j-1}^2)q^{k-1}(\sigma_{j+1}^2)\cdots q^{k-1}(\sigma_{N_C}^2)q^{k-1}(\sigma_{\mathbf{c}}^2)}\left[\ln p(\mathbf{c},\sigma_1^2,\sigma_2^2,...,\sigma_{N_C}^2,\sigma_{\mathbf{c}}^2,\mathbf{R})\right]\\
&=\mathbb{E}_{q^k(\mathbf{c})q^k(\sigma_1^2)\cdots q^k(\sigma_{j-1}^2)q^{k-1}(\sigma_{j+1}^2)\cdots q^{k-1}(\sigma_{N_C}^2)q^{k-1}(\sigma_{\mathbf{c}}^2)}\Big[(-\alpha'_j-1)\ln\sigma_j^2-\frac{\beta_j}{\sigma_j^2}\\
&\qquad -\frac{1}{2\sigma_j^2}(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})^T(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})+\zeta(\mathbf{c},\sigma_1^2,...,\sigma_{j-1}^2,\sigma_{j+1}^2,...,\sigma_{N_C}^2,\sigma_{\mathbf{c}}^2,\mathbf{R})\Big]\\
&=(-\alpha'_j-1)\ln\sigma_j^2-\frac{\beta_j}{\sigma_j^2}-\frac{1}{2\sigma_j^2}\mathbb{E}_{q^k(\mathbf{c})}\left[(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})^T(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})\right]\\
&\qquad\qquad +\mathbb{E}_{q^k(\mathbf{c})q^k(\sigma_1^2)\cdots q^k(\sigma_{j-1}^2)q^{k-1}(\sigma_{j+1}^2)\cdots q^{k-1}(\sigma_{N_C}^2)q^{k-1}(\sigma_{\mathbf{c}}^2)}\\
&\qquad\qquad\qquad \left[\zeta(\mathbf{c},\sigma_1^2,...,\sigma_{j-1}^2,\sigma_{j+1}^2,...,\sigma_{N_C}^2,\sigma_{\mathbf{c}}^2,\mathbf{R})\right],
\end{aligned}$$

where $\zeta(\mathbf{c},\sigma_1^2,...,\sigma_{j-1}^2,\sigma_{j+1}^2,...,\sigma_{N_C}^2,\sigma_{\mathbf{c}}^2,\mathbf{R})$ contains terms in $\ln p(\mathbf{c},\sigma_1^2,...,\sigma_{N_C}^2,\sigma_{\mathbf{c}}^2,\mathbf{R})$ not involving σ_j. Taking into account the normalization condition for $q^k(\sigma_j^2)$ and comparing the expression with the defining equation of the inverse Gamma distribution, we conclude that for $j=1,...,N_C$

$$q^k(\sigma_j^2)=IG\left(\sigma_j^2;\alpha'_j,\beta_j+\frac{1}{2}\mathbb{E}_{q^k(\mathbf{c})}\left[(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})^T(\mathbf{R}^j-\mathbf{K}^j\mathbf{c})\right]\right).$$

Finally, by taking the variational derivative of the Lagrange function $\mathfrak{L}\left(q(\mathbf{c}),q(\sigma_1^2),\cdots,q(\sigma_{N_C}^2),q(\sigma_{\mathbf{c}}^2)\right)$ with respect to $q(\sigma_{\mathbf{c}}^2)$ and equating it to zero, we can deduce that

$$q^k(\sigma_{\mathbf{c}}^2)=IG\left(\sigma_{\mathbf{c}}^2;\alpha_{\mathbf{c}}+\frac{n}{2},\beta_{\mathbf{c}}+\frac{1}{2}\mathbb{E}_{q^k(\mathbf{c})}\left[(\mathbf{Lc})^T(\mathbf{Lc})\right]\right).$$

2. Finite element approximation of rate constant map. Since the solution to our a single step model, an integral equation

$$\int_\Omega K(t,C;k_a,k_d)f(k_a,k_d)dk_adk_d=R_{obs}(t;C), \tag{2}$$

of estimating the numbers and rate constants of interactions in a chemical reaction does not have an analytically closed form, we need to solve our continuous model numerically. Moreover, the representation of rate constant map in computer also require an appropriate discretization of a function. In this work, we employee the finite element method due to its geometric flexibility.

Following Johnson (2009); Zhang et al. (2017), we discretize our bounded domain $\Omega\subset\mathbb{R}^2$ by mesh $\mathcal{T}$ using non-overlapping triangles $\{\triangle_\mu\}_{\mu=1}^M$, i.e., $\Omega=\sqcup_{\triangle_\mu\in\mathcal{T}}\triangle=\sqcup_{\mu=1}^M\triangle_\mu$. We associate the triangulation $\mathcal{T}$ with the mesh function $h(x)$, which is a piecewise-constant function such that $h(x) \equiv \ell(\triangle_\mu)$ for all $x \in \triangle_\mu \in \mathcal{T}$, where $\ell(\triangle_\mu)$ is the longest side of $\triangle_\mu$. Define the mesh scale size as $h := \max_{x\in\Omega} h(x) \equiv \max_\mu \ell(\triangle_\mu)$. Let $r(\triangle_\mu)$ be the radius of the maximal circle contained in the triangle $\triangle_\mu$. We make the following shape regularity assumption for every element $\triangle_\mu \in \mathcal{T}$: $c_1 \leq \ell(\triangle_\mu) \leq c_2 r(\triangle_\mu)$, where c_1 and c_2 are two positive constants.

Now, we introduce the finite element space

$$V_n = \left\{ f \in L^2(\Omega) : \ f \in \mathcal{P}_1(\triangle_\mu) \text{ for all } \triangle_\mu \in \mathcal{T} \right\}, \tag{3}$$

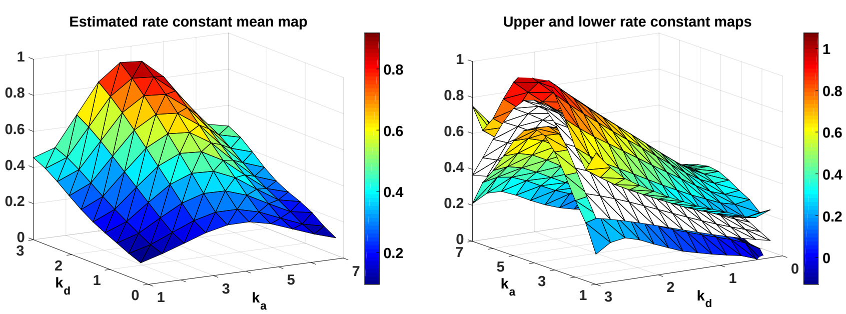
where $L^2(\Omega)$ presents the space of square integrable functions defined on Ω, and $\mathcal{P}_1(\triangle_\mu)$ denotes the set of all piecewise-linear functions on $\triangle_\mu$.

Theorem S1 *Let f and f_n be the solutions of integral equation (2) in $H^1(\Omega)$ and V_n respectively. Then, we have*

$$\|f_n - f\|_{L^2(\Omega)} \leq Ch \|\nabla f\|_{L^2(\Omega)}, \tag{4}$$

where C is a constant, depending only on the domain Ω and triangulation $\mathcal{T}$. In other words, $f_n \to f$ in $L^2(\Omega)$ as $h \to 0$.

3. A demonstration of Algorithm 2. This section presents the working procedure at the main steps of Algorithm 2. Assume that we have the measurement with small noise $\delta = 0.001$. At the initialization step of the algorithm, the domain Ω is discretized using 162 triangular finite elements



(a) (b)

FIG 1. *(a) The estimated rate constant mean map at Step 3 of Algorithm 2, i.e. the result of Algorithm 1. (b) The estimated lower and upper rate constant maps at Step 4 of Algorithm 2. The white surface denotes the exact rate constant map, which is located between the lower and upper rate constant maps.*

with 100 vertices. The pairs of parameters $\{(\alpha_j, \alpha_j)\}_{j=1}^{N_C}$ and $(\alpha_{\mathbf{c}}, \alpha_{\mathbf{c}})$ for the inverse Gamma distribution are all taken to be $(1, 1)$. The initial guess of the distribution of the rate constant map is multi-dimensional normal distribution with mean $\mathbf{c}^* = \left(\mathbf{K}^T\mathbf{K}\right)^+ \mathbf{K}^T\mathbf{R}$ and identity covariance matrix $\mathbf{I}$, where the superscript "+" denotes the Moore-Penrose inversion. Step 3 of Algorithm 2, i.e. Algorithm 1, gives (a) in Figure 1. The computed lower and upper rate constant maps at Step 4 of Algorithm 2 are displayed in (b) of Figure 1. The evolution of Algorithm 2 can be found in Figure 2.

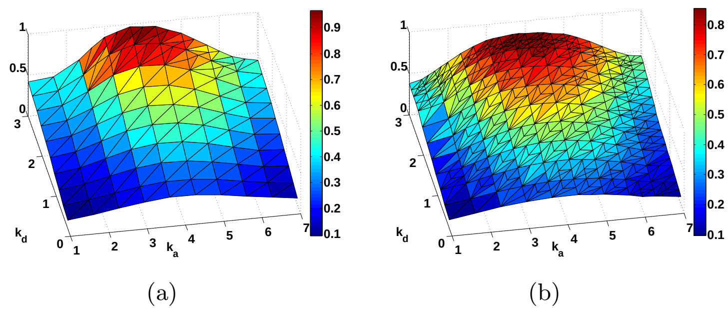


FIG 2. *The evolution of Algorithm 2: (a) After one iteration. The number of nodes equals 124, and the number of triangles is 203. L2Err=0.6241. (b) After five iterations. The number of nodes equals 417, and the number of triangles is 766. L2Err=0.2530.*

References.

JOHNSON, C. (2009). *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Mineola: Dover.

ZHANG, Y., YAO, Z., FORSSÉN, P. and FORNSTEDT, T. Estimating the Rate Constant from Biosensor Data via an Adaptive Variational Bayesian Approach. *submitted to Ann. Appl. Stat.*

ZHANG, Y., FORSSÉN, P., FORNSTEDT, T., GULLIKSSON, M. and DAI, X. (2017). An adaptive regularization algorithm for recovering the rate constant distribution from biosensor data. *Inverse Problems in Science & Engineering* **DOI: 10.1080/17415977.2017.1411912**.

Y. ZHANG
FACULTY OF MATHEMATICS
CHEMNITZ UNIVERSITY OF TECHNOLOGY
09107 CHEMNITZ, GERMANY
E-MAIL: ye.zhang@oru.se

Z. YAO
CORRESPONDING AUTHOR
DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
21 LOWER KENT RIDGE ROAD
SINGAPORE 117546
E-MAIL: zhigang.yao@nus.edu.sg

P. FORSSÉN AND T. FORNSTEDT
DEPARTMENT OF ENGINEERING AND CHEMICAL SCIENCES
KARLSTAD UNIVERSITY
65188 KARLSTAD, SWEDEN
E-MAIL: patrik.forssen@kau.se
E-MAIL: torgny.fornstedt@kau.se