ESTIMATING THE RATE CONSTANT FROM BIOSENSOR DATA VIA AN ADAPTIVE VARIATIONAL BAYESIAN APPROACH

BY YE ZHANG[*,1], ZHIGANG YAO[†,2,4], PATRIK FORSSÉN[‡,3] AND TORGNY FORNSTEDT[‡,3]

Chemnitz University of Technology[*], *National University of Singapore*[†] *and Karlstad University*[‡]

The means to obtain the rate constants of a chemical reaction is a fundamental open problem in both science and the industry. Traditional techniques for finding rate constants require either chemical modifications of the reactants or indirect measurements. The rate constant map method is a modern technique to study binding equilibrium and kinetics in chemical reactions. Finding a rate constant map from biosensor data is an ill-posed inverse problem that is usually solved by regularization. In this work, rather than finding a deterministic regularized rate constant map that does not provide uncertainty quantification of the solution, we develop an adaptive variational Bayesian approach to estimate the distribution of the rate constant map, from which some intrinsic properties of a chemical reaction can be explored, including information about rate constants. Our new approach is more realistic than the existing approaches used for biosensors and allows us to estimate the dynamics of the interactions, which are usually hidden in a deterministic approximate solution. We verify the performance of the new proposed method by numerical simulations, and compare it with the Markov chain Monte Carlo algorithm. The results illustrate that the variational method can reliably capture the posterior distribution in a computationally efficient way. Finally, the developed method is also tested on the real biosensor data (parathyroid hormone), where we provide two novel analysis tools—the thresholding contour map and the high order moment map—to estimate the number of interactions as well as their rate constants.

1. Introduction. In the modern world, biosensors have made a significant impact in many fields, such as antibody-antigen interactions, immunology, virology and the pharmaceutical industry (Elmlund et al. (2015), Sanvicens et al. (2011)).
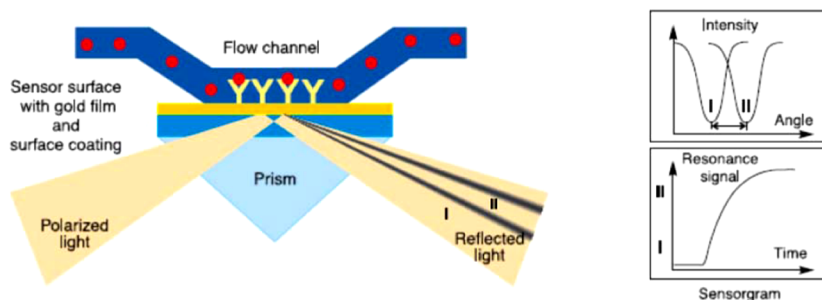
Fig. 1. *A simplified biosensor system.*

Hence, during the last few decades, there has been an accelerated technological development of biosensor instruments, for example, surface plasmon resonance, quartz crystal microbalance, etc. A simplified biosensor system is presented in Figure 1, and its physical mechanism is briefly discussed in Appendix A. To design an appropriate biosensor instrument that is biocompatible or specifically functionalized, scientists must know the physical chemistry of biomolecular/cell surface interactions (Telesca et al. (2012)). The reliable analysis of biomolecular interactions is crucial in both science and the industry, for example, it is required to fulfill modern drug quality assurance criteria. In order to understand the interactions, scientists collect biosensor data that measures the analyte biomolecules of several different concentrations on a sensor chip with immobilized ligand molecules that form complexes with analytes. This kind of biosensor data is usually called a sensorgram, where the systems response, proportional to total complex concentration, is measured over time for different analyte injections. In this paper, we focus on this particular type of data and aim to obtain information about the interactions, that is, their numbers and the corresponding rate constants. The biosensor data is traditionally processed using a simple model fitting procedure assuming just one, or perhaps two, distinct interactions. These traditional approaches might not reflect the true, complicated, and heterogeneous molecular interactions for large active pharmaceutical ingredients, and can thus lead to wrong mechanistic conclusions. Therefore, a more advanced analysis of the biosensor data is necessary to avoid the costly and time consuming procedure of repeating the same ligand-analyte binding experiment over and over using chips with different ligand binding or different buffers, etc., until one of the tests results in "good enough" data that works with the standard rate constant estimation (Gray and Pierce (1985)).

1.1. *Existing work in statistics and computational chemistry.* The problem of estimating the chemical rate constants of interactions through statistical approaches has been investigated for long time (Atherton, Schainker and Ducot (1975), Box and Draper (1965), Gupta and Rawlings (2014), Komorowski et al. (2011), Pischel, Sundmacher and Flassig (2017), Ziegel and Gorman (1980)). To

our best knowledge, most of the existing statistical approaches require a priori knowledge of the number of interactions in a chemical reaction, that is, they first assume the existence of parallel reaction (due to potential interactions), and then estimate the rate constants for corresponding interactions. Therefore, in practice, the problem is usually solved by a two-step strategy. The first step is to determine the number of interactions by dimensionality reduction technique (e.g., principal component analysis, regression tree (Loh and Zheng (2013))), and then fit the data to the parameters (i.e., rate constants correspond to these reactions). It is clear that the first step partially determines the quality of the results in the second step. However, it can be observed that the second step procedure also potentially has a positive influence on the estimation in the first step, particularly when the data is under-sampled and affected by large noise. Therefore, the separation of the procedure into two steps does not seem optimal. The question that now arises is whether the number of interactions and the values of rate constants corresponding to each interaction in a single step can be estimated from the biosensor data, therefore making the modeling more robust. These considerations form the motivations behind our work.

Another line of research, inspired by Adsorption Energy Distribution (AED) calculations (Stanley and Guiochon (1995)), which are successfully used for steady-state data usually from small molecular systems, as well as a couple of numerical solvers for dynamic biosensor data, such as the Interaction Distribution (ID) (or Interaction Map) (Altschuh et al. (2012), Svitel et al. (2003)) and Adaptive Interaction Distribution Algorithm (AIDA) (Zhang et al. (2018)), have recently gained much attention. With these solvers, one can see the number of interactions in the system, along with their associated rate constants, as peaks on a surface. The mathematical basis behind these solvers is to use a single step model to estimate the number of rate constants and their values simultaneously. They utilize regularization methods to impose a sense of well-posedness for the model and solve the resulting deterministic problems by optimization. However, the result of these methods strongly depends on the choice of system parameters, for example, the type of regularization penalty and the value of the regularization parameter. For our application, as well as other problems where no ground truth is available, the problem of selecting the regularization penalty term and the regularization parameter remains an open issue (Fan and Lv (2008)), as most tuning parameters such as the generalized cross-validation (GCV) (Tibshirani (1996)), C_p style statistics (Efron et al. (2004)), and the consistent estimator of degrees of freedom of the LASSO in the C_p, AIC, and BIC criteria (Zou, Hastie and Tibshirani (2007)), are based on an asymptotic property that is practically infeasible.

The shortage of existing methods motivates us to combine statistical and deterministic methods to develop a more robust single step approach via a statistical strategy. Note that this high level idea has been used in Gorshkova et al. (2008) to some extent. However, from a mathematical viewpoint, their method can still be classified in the branch of deterministic methods, since it estimates the distribution

of rate constant directly by solving a special regularized optimization problem. Though the authors have taken into account the contribution of a prior expectation of rate constants on the regularization term in their optimization formulation, the mathematical model should be deterministic if one considers the distribution of rate constant, which is also termed as the rate constant map, as a deterministic object (i.e., an element in the space of functions). Furthermore, the justification of rate constants as random variables is missing, and no uncertainty quantification is discussed under their framework. Meanwhile, the difficulty of choosing an appropriate regularization parameter remains. This motivates us to regard the rate constant map as a random variable, which incorporates the regularization in a completely different way, and overcomes the difficulty of selecting regularization parameter, hence may carry the intrinsic information from the uncertainty quantification of the rate constant map.

1.2. *Motivation.* As discussed above, due to the diversity of data structure and the appearance of noise in many biosensor problems, the existing methods do not work efficiently. The central problem is that most of them are not capable of inferring the interaction information correctly. For instance, in the assumption that two interactions exist in our biosensor system, the exact value of rate constants in our data structure called the dissociation constant k_d (measures the rate at which a molecule complex diassociates) and association constant k_a (rate at which molecules form complexes) (see Appendix A for details) are displayed in (a) of Figure 2. This means that two parallel interactions exist and their corresponding rate constants are $(\log_{10}(k_d), \log_{10}(k_a)) = (-3, 4.5)$ and $(-2, 6.5)$. Now we solve
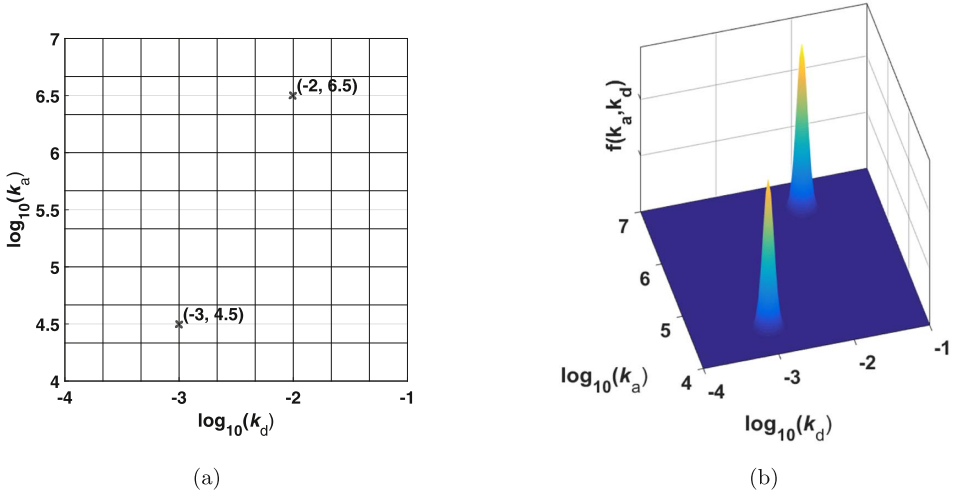

(a)   (b)

FIG. 2. (a) *The exact value of association and dissociation constants.* (b) *One example of the expected rate constant map* $f(k_a, k_d)$.
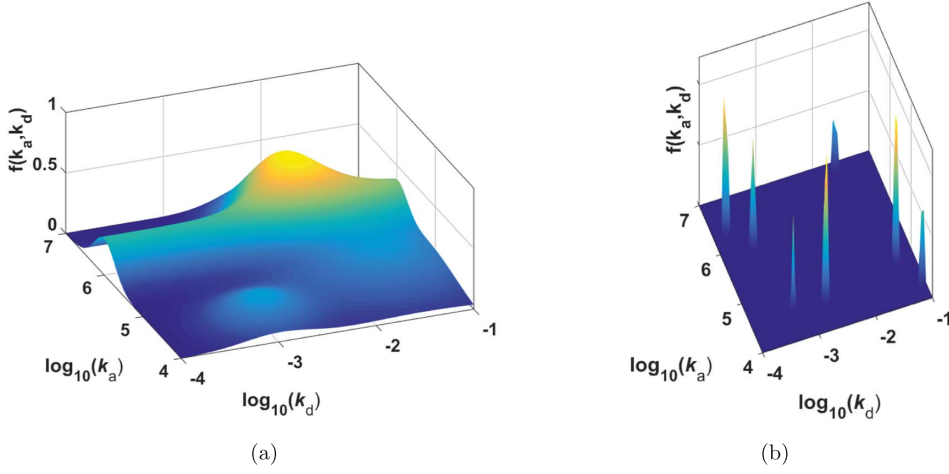
(a)

(b)

FIG. 3. (*a*) *The Tikhonov method. The estimated rate constant map is too smooth to enable the deduction of the number of interactions and the value of association and dissociation constants (the position of peaks in the rate constant map). (b) The sparse regularization method (ℓ_1 regularization). The information of the interaction is completely hidden in the estimated rate constant map.*

the problem using the conventional rate constant map method, which will be introduced later in Section 2 and Appendix A. This method provides a map of distribution of rate constants (called a rate constant map), from which we can figure out the number of interactions and the rate constants of these interactions. The desired estimated rate constant map should exhibit the information of the interactions. One example of this rate constant map is displayed in (b) of Figure 2, where people can easily derive the interaction numbers and rate constants from the peaks of the rate constant map. However, in many cases, the regularized deterministic approaches cannot offer the rate constant map that is correctly shaped. For instance, in our parathyroid hormone application, the Tikhonov regularization provides an oversmoothing effect on the rate constant map, while, the sparse ℓ_1 regularization offers a rather unsatisfactory reconstruction of the rate constant map; see Figure 3 for details. This motivates us to build new statistical models and investigate the methods that can efficiently estimate the rate constant map with uncertainty quantification, from which we can explore the intrinsic heterogeneity of the biosensor system—the interaction information including the number of interactions and the corresponding rate constants. The philosophy of our approach is to assume that the rate constants are no longer static. By utilizing the nonstatic rate constants, where the rate constant map f is modeled as a high dimensional random variable, we will be able to investigate the distribution of f and provide estimates of the variability. Based on the estimated distribution of f, we can produce different kinds of maps (e.g., (b) in Figure 2 or (a) and (b) in Figure 3) which provides a multi-scale perspective of the rate constant map. It turns out that the random rate constant map f

can be used to find the interaction information, for example, the number of the interactions and rate constants for each interaction, which will be called the intrinsic quality of a chemical system.

In comparison with classical deterministic regularization methods, Bayesian approaches have distinct features for inverse problems (in our case, the inverse problem is the estimation of interaction information from indirectly measured biosensor data) that make them attractive for statistical inference (Bernardo and Smith (1994), Evans and Stark (2002), Kaipio and Somersalo (2007), Kennedy and O'Hagan (2001)). First, the Bayesian posterior distribution provides uncertainty quantification by assigning probabilities to the space of all possible inverse solutions that are consistent with the observed data. This uncertainty quantification is lacking in all the deterministic regularization methods, which only yield one single solution as a point estimate. Second, the use of prior distribution allows more flexibility in modeling. For instance, sparsity-inducing priors such as the spike-and-slab prior can be used if sparse solutions are desirable for high dimensional models. Constraints on the parameter space, or the space of solutions in our context, can be imposed directly through constraining the support of the prior distribution. Third, the tuning parameters that are used in the frequentist regularization methods become hyper-parameters in the Bayesian framework. They are usually assigned some hyper-priors and can hence be made fully adaptive to the observed data. On the other hand, since the posterior distribution does not have an analytically tractable closed form in most cases, Bayesian inference typically requires the numerical sampling of the posterior distributions by Markov chain Monte Carlo (MCMC), which can be computationally intensive and may suffer from slow mixing due to high dimensionality or strong correlations between the parameters in their joint posterior distribution. As an alternative approach, the variational Bayesian (VB) inference (Bishop (2006), Blei, Kucukelbir and McAuliffe (2017)) has gained popularity for its computational efficiency over posterior sampling methods in such complex Bayesian models. VB approximates the true posterior distribution with a family of computationally tractable distributions by minimizing the KL divergence between them. Although it has been empirically observed that VB occasionally underestimates the posterior variances, the gain in computational efficiency from VB is substantial in many applications; see, for example, Parisi (1988), Peterson and Anderson (1987), Jordan et al. (1999), Wainwright and Jordan (2008), Jin and Zou (2010) and references therein.

In this paper, we build a Bayesian framework for uncertainty quantification in the biosensor problem and propose a new VB approach for computing the posterior distribution. Our VB approach is novel and motivated by three physical facts that are unique to the biosensor problem, as well as other similar problems, modeled by the integral equations (1). First, in our real world biosensor problem, the noise structure can be heterogeneous among different sensorgrams, which results in extremely high dimensionality of the parameter spaces that makes any posterior sampling algorithms such as the MCMC practically infeasible. Therefore, we develop

a VB approach to circumvent the computational problem. Second, the solution of the rate constant map must be nonnegative for meaningful physical interpretation. This fact requires a constrained prior support and eventually a new VB algorithm that differs from the existing ones. Finally, the dimensionality of parameter space in most existing statistical models should be fixed, which requires the a priori knowledge of the distribution of rate constants in the biosensor system that is actually unknown before we estimate them. To overcome this difficulty, based on the newly developed VB algorithm we introduce an adaptive discretization technique that automatically adjusts the dimensionality of solution space during the evolution of our algorithm, according to the uncertainty quantification of the obtained temporary qualities.

1.3. *Objective of the paper.* In Section 2, based on the rate constant map theory, we establish a single step model connecting the biosensor data and the rate constant map, which is used for estimating the number of interactions and the rate constants simultaneously. Section 3 presents our newly developed method—an Adaptive Variational Bayesian Approach (AVBA)—for solving the proposed single step model. In Section 4, we illustrate the approach using an artificial problem, and compare its performance with the MCMC algorithm. The proposed method is tested on real biosensor data (parathyroid hormone) in Section 5, where we will develop two AVBA-based analysis tools that represent the intrinsic property of the interaction, and can be used for estimating the number of interactions as well as their rate constants. Finally, concluding remarks are given in Section 6.

2. A mathematical model for estimating the rate constant map. As mentioned before, in this work, we use the rate constant map $f(k_a, k_d)$ (i.e., the distribution of association constant k_a and dissociation constant k_d) to explore the interaction information, that is, their numbers and (active) rate constants. Based on the *Rate Constant Map Theory* (see details in the Appendix A), the measured sensorgram R_{obs}, which is dependent on the analyte concentration C and time t, and the rate constant map f are related through the following integral equation

$$(1) \qquad R_{\text{obs}}(t; C) = \int_{\Omega} K(t, C; k_a, k_d) f(k_a, k_d) \, dk_a \, dk_d, \quad (k_a, k_d) \in \Omega,$$

where $\Omega \subset \mathbb{R}^2$ is the interested domain of rate constants, and the kernel function $K(\cdot)$ is defined as

$$(2) \qquad K(t, C; k_a, k_d) = \begin{cases} 0 & t \le t_0 + \Delta t, \\ \dfrac{k_a C}{k_d + k_a C}\left(1 - e^{-(k_d + k_a C)(t - t_0)}\right), \\ \quad t_0 + \Delta t < t \le t_0 + t_{inj} + \Delta t, \\ \dfrac{k_a C}{k_d + k_a C}\left(1 - e^{-(k_d + k_a C)t_{inj}}\right)e^{-k_d(t - t_0 - t_{inj})}, \\ \quad t > t_0 + t_{inj} + \Delta t, \end{cases}$$

where the physical meaning of parameters t_0, t_{inj}, Δt can be found in Appendix A. All of them are given constants.

Since the solution to the integral equation (1) does not have an analytically closed form, we have to solve (1) numerically. Moreover, in order to represent the rate constant map on computer, we need to discretize the function $f(k_a, k_d)$ by a vector $\mathbf{c}$, associated with appropriate coordinates. By slightly abusing the notation, we shall call $\mathbf{c}$ the rate constant map as well. In this work, we adapt the finite element technique for discretization. The relation between f and its finite element approximation is provided in the Supplementary Material (Zhang et al. (2019)). We discretize our bounded domain $\Omega \subset \mathbb{R}^2$ by mesh $\mathcal{T}$ using nonoverlapping triangles $\{\triangle_\mu\}_{\mu=1}^M$ with the standard shape regularity assumption. Introduce the finite element space V_n of piecewise linear elements on the triangulation $\mathcal{T}$ with the number of notes n. Then, the projection of $f(k_a, k_d)$ in V_n (denote as f_n) can be decomposed as

$$(3) \qquad f_n(k_a, k_d) = \sum_{l=1}^n \mathbf{c}_l \cdot \varphi_l(k_a, k_d),$$

where $\{\mathbf{c}_l\}$ are the coordinates of f_n under basis $\{\varphi_l\}$ of the linear space V_n.

REMARK 1. The dimensionality n of approximate solution space cannot be fixed, since we have no a priori information about the distribution of peaks of the rate constant map f, and cannot provide a good mesh $\mathcal{T}_n$ which nodes hit or close to the positions of the peaks of f. A naive way to capture the locations of the peaks of f is to use an intensive grid with large n so that every possible peak is located in a small neighborhood of a node of mesh. However, such an idea cannot be implemented in practice due to computational complexity. Alternatively, in this work, we start with a small n paired with a coarse mesh $\mathcal{T}_n$, and then increase n with a finer local mesh that only covers the region with possible peaks of f until an approximate solution with satisfactory accuracy is obtained. Hence, instead of a fixed model (5) below, our algorithm will deal with a sequence of models of type (5) with dynamic dimensionality n.

Furthermore, denote by $\mathbf{R}^j = [R_{\mathrm{obs}}(t_1; C_j), R_{\mathrm{obs}}(t_2; C_j), \ldots, R_{\mathrm{obs}}(t_{N_T}; C_j)]^T$ the measured sensorgram with the injection C_j and the time grid $\{t_i\}_{i=1}^{N_T}$. Then, using the decomposition (3) and considering the inaccuracy of the measurement data, the finite element approximation of integral equation (1) can be written as a system of algebraic equations:

$$(4) \qquad \mathbf{R}^j = \mathbf{K}^j \mathbf{c} + \boldsymbol{\epsilon}^j, \quad j = 1, \ldots, N_C,$$

where $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n]^T$, $\boldsymbol{\epsilon}^j \in \mathbb{R}^{N_T}$ is the additive error for jth sensorgram, and matrix $\mathbf{K}^j$ is defined by $[\mathbf{K}^j]_{i,l} = \int_\Omega K(t_i, C_j; k_a, k_d) \varphi_l(k_a, k_d) \, dk_a \, dk_d$.

Denote $\mathbf{R} = [\mathbf{R}^1; \ldots; \mathbf{R}^{N_C}]$, $\mathbf{K} = [\mathbf{K}^1; \ldots; \mathbf{K}^{N_C}]$ and $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}^1; \ldots; \boldsymbol{\epsilon}^{N_C}]$, where for column vectors $\mathbf{a}$ and $\mathbf{b}$, $[\mathbf{a}; \mathbf{b}] = [\mathbf{a}^T, \mathbf{b}^T]^T$. Then, the system (4) can be rewritten into a compact form

$$(5) \qquad\qquad\qquad \mathbf{R} = \mathbf{Kc} + \boldsymbol{\epsilon}.$$

The deterministic approaches solve (5) by the iterative regularization (e.g., the Landweber method Engl, Hanke and Neubauer (1996)), or the variational regularization with appropriate regularization term (e.g., the Tikhonov method), to obtain a regularized approximate solution $\mathbf{c}_{\text{reg}}$ of (5). Using the obtained $\mathbf{c}_{\text{reg}}$, an approximate rate constant map f_n can be reconstructed by (3), which is a function well defined in the whole domain Ω. The active association and dissociation constants (k_a and k_d) for the given kinetic reaction process of the biomolecular system should be the peaks in the estimated rate constant map f_n; see Svitel et al. (2003) and Altschuh et al. (2012). In this work, we refer to these peaks as the Intrinsic Property of the Interaction (IPoI). Now the problem at hand is how to determine the IPoI in a chemical reaction through the single step model (1) (or its finite dimensional analogue (5)). This problem is seriously ill-posed since (i) one can always construct a large number of (deterministic) rate constant map f satisfying the equation (1). In the reduced finite dimensional model (5), various numerical experiments have shown that rank($\mathbf{K}_{N_T N_C \times n}) \ll \min\{N_T N_C, n\}$. (ii) Note that the integral operator in (1) is a compact operator, which is affected by an instability phenomenon where a small amount of noise in the measurement data can lead to enormous errors in the estimates (Tikhonov and Arsenin (1977)). Moreover, solutions to the deterministic models, for example, the regularized solutions in Svitel et al. (2003), Altschuh et al. (2012), Gorshkova et al. (2008) and Zhang et al. (2018), do not represent the pure information of IPoI even though the experimental data is accurate enough. This phenomenon implies that the deterministic regularized solutions contain redundant information about a chemical reaction besides the IPoI. This drawback, as well as the problem of choosing a regularization term in the deterministic models aforementioned, motivates us to model the quantity $\mathbf{c}$ in (5) (or f in (1)) as a random vector (or function), from which we can study the uncertainty of the rate constant. As demonstrated in Section 5, the uncertainty of the quantity $\mathbf{c}$ can be used to interpret the difference between the deterministic solution of (5) and the IPoI.

3. An adaptive variational Bayesian approach.

3.1. *Bayesian framework.* A detailed understanding of the dynamics of the rate constant map $\mathbf{c}$ given data $\mathbf{R}$ and the model (5) is crucial in the context of our problem.

For simplicity, a Gaussian distribution of noise on each sensorgram is customarily assumed, that is, $\boldsymbol{\epsilon}^j$ ($j = 1, \ldots, N_C$) are independent and identically distributed (i.i.d.) additive Gaussian random vectors with mean zero and variance

$\sigma_j^2 I_{N_T}$, where I_{N_T} denotes the identity matrix of size N_T. The assumption of normality is preferred due to the fact that Gaussian noise is present at the sensorgrams, and sensorgram noise is usually much smaller than the signal. Although correlated sensorgram noise is more realistic than "homogenerous" sensorgrams, this could complicate the problem. Then, the likelihood $p(\mathbf{R}|\mathbf{c})$ is given by

$$(6) \qquad p(\mathbf{R}|\mathbf{c}) \propto \exp\left\{-\frac{1}{2}(\mathbf{R} - \mathbf{Kc})^T \Sigma^{-1}(\mathbf{R} - \mathbf{Kc})\right\} \prod_{j=1}^{N_C}\{\sigma_j^{-N_T}\},$$

where $\propto$ denotes up to a multiplicative normalizing constant, and the covariance matrix of noise $\Sigma = \mathrm{diag}(\sigma_1^2 I_{N_T}, \ldots, \sigma_{N_C}^2 I_{N_T})$.

A natural mechanism for regularization is in the form of *prior* information. A versatile prior distribution of $\mathbf{c}$ is the Markov random field prior; see Babacan, Molina and Katsaggelos (2008) for details. Define $\mathbf{1}_\Theta$ as the indicator function on the set Θ, that is, $\mathbf{1}_\Theta(\mathbf{c}) = 1$ for $\mathbf{c} \in \Theta$, while $\mathbf{1}_\Theta(\mathbf{c}) = 0$ for $\mathbf{c} \notin \Theta$. Denote by $\mathbf{c} \geq 0$ the nonnegativity by components. For computational tractability, we use the following conjugate prior distribution

$$(7) \qquad p(\mathbf{c}) \propto \sigma_{\mathbf{c}}^{-n} \exp\left\{-\frac{1}{2\sigma_{\mathbf{c}}^2}(\mathbf{Lc})^T(\mathbf{Lc})\right\}\mathbf{1}_{\mathbf{c}\geq 0},$$

where the matrix $\mathbf{L}$ encapsulates the structure of the interactions between neighboring components of the solution $\mathbf{c}$. $\mathbf{L}$ is given and assumed to be full rank. Note that our choice of (7) also reflects the physical meaning of the rate constant. If the hyper-parameters, $\{\sigma_j\}_{j=1}^{N_C}$ and $\sigma_{\mathbf{c}}^2$, are given and the nonnegative constraint of the solution $\mathbf{c}$ is ignored, the prior of $\mathbf{c}$ in (7) for the *maximum a posteriori* estimate is exactly the same as the L_2-type penalty term $\|\mathbf{Lc}\|^2$ in the classical variational regularization theory (Stuart (2010)), where $\|\cdot\|$ denotes the standard Euclidean norm of a vector or the Frobenius norm of a matrix. For the classical Tikhonov regularization, $\mathbf{L}$ stands for the discrete Laplacian.

To complete the model specification, we impose conjugate priors on the parameters $\{\sigma_j^2\}_{j=1}^{N_C}$ and $\sigma_{\mathbf{c}}^2$. In other words, for each of these parameters, we impose an independent inverse-gamma prior with the density $\mathrm{IG}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp(-\frac{\beta}{x})$, where α is the shape parameter, β is the scale parameter, and $\Gamma(\cdot)$ denotes the standard Gamma function. Then, plugging-in the density functions in (5), the posterior distribution of $\mathbf{c}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2$ given the data $\mathbf{R}$ is

$$p(\mathbf{c}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2 | \mathbf{R})$$

$$(8) \qquad \propto p(\mathbf{R}|\mathbf{c}, \sigma_{\mathbf{c}}^2) \cdot p(\mathbf{c}|\sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_C}^2) \cdot p(\sigma_{\mathbf{c}}^2) \cdot \prod_{j=1}^{N_C} p(\sigma_j^2)$$

$$\propto \sigma_{\mathbf{c}}^{-n-2\alpha_{\mathbf{c}}-2} \cdot \exp\left(-\frac{\beta_{\mathbf{c}}}{\sigma_{\mathbf{c}}^2}\right) \prod_{j=1}^{N_C} \left\{\sigma_j^{-N_T-2\alpha_j-2} \exp\left(-\frac{\beta_j}{\sigma_j^2}\right)\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}(\mathbf{R} - \mathbf{Kc})^T \Sigma^{-1}(\mathbf{R} - \mathbf{Kc}) - \frac{1}{2\sigma_{\mathbf{c}}^2}(\mathbf{Lc})^T (\mathbf{Lc})\right\} \cdot \mathbf{1}_{\mathbf{c}\geq 0},$$

where the values of hyperparameters $(\alpha_{\mathbf{c}}, \beta_{\mathbf{c}})$ and (α_j, β_j) are given based on simulation results of artificial problems. The posterior distribution $p(\mathbf{c}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2 | \mathbf{R})$ in (8) is the full Bayesian solution to our finite model (5), and it encapsulates all the information including IPoI. However, due to the high dimensionality of $\mathbf{c}$, (8) is also a distribution that lives in a space of very high dimension, for example, $10^3 \sim 10^5$, for our real data. Hence, it is advisable to develop tools for exploring this very high dimensional posterior distribution. In the next section, we develop an approximate inference method based on the mean field variational approximation for exploring the posterior in (8).

3.2. *Variational approximation algorithm.* Due to the presence of several variance parameters, the posterior distribution (8) does not have an explicit closed form. One way to explore this high dimensional posterior is to perform posterior sampling based on MCMC algorithms. Although the MCMC has the advantage of being asymptotically exact (Robert and Casella (2004)), its convergence is also often known to be difficult to diagnose (Bishop (2006), Brooks et al. (2002)), and it suffers from slow mixing in the presence of high dimensional parameters. As the dimensionality of the problem in the biosensor system is extremely large, we shall take an alternative route and focus on the variational approximation approach. The essential idea of variational methods consists of first transforming the problem into an equivalent optimization problem, and then obtaining an approximate distribution $q(\cdot)$ to the true posterior $p(\cdot)$ by solving the optimization problem inexactly. The idea behind VB approximation is to find a family of tractable distributions $\{q^k(\cdot)\}_k$ to approximate the posterior density by minimizing the Kullback–Leibler (KL) divergence (Kullback (1959)), while still capturing distinct features of the posterior distribution (8) in a computationally efficient way.

The KL divergence is a nonsymmetric measure of the difference between two probability distributions, and is defined as

$$\mathcal{D}\big(q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2) | p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2 | \mathbf{R})\big)$$

$$= \int q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)$$

$$\times \ln\left(\frac{q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)}{p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2 | \mathbf{R})}\right) d\mathbf{c} \, d\sigma_1^2 \cdots d\sigma_{N_C}^2 \, d\sigma_{\mathbf{c}}^2$$

(9)

$$= \int q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)$$

$$\times \ln\left(\frac{q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)}{p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})}\right) d\mathbf{c} \, d\sigma_1^2 \cdots d\sigma_{N_C}^2 \, d\sigma_{\mathbf{c}}^2$$

$$+ \ln(p(\mathbf{R})) := \mathrm{ELBO}(q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2) | p(\bullet, \mathbf{R})) + \ln(p(\mathbf{R})),$$

where $p(\mathbf{R}) = \int p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R}) \, d\mathbf{c} \, d\sigma_1^2 \cdots d\sigma_{N_C}^2 \, d\sigma_{\mathbf{c}}^2$ is a normalizing constant, and the global distribution $p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$ has the form

(10)
$$p(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2, \mathbf{R})$$
$$= p(\mathbf{c}|\sigma_{\mathbf{c}}^2) \cdot p(\mathbf{R}|\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2) \cdot p(\sigma_{\mathbf{c}}^2) \cdot \prod_{j=1}^{N_C} p(\sigma_j^2).$$

Note that minimizing the KL distance in (9) is equivalent to minimizing the first term, which is termed as the *evidence lower bound* and denoted by ELBO (Blei, Kucukelbir and McAuliffe (2017)). In this way, we have successfully transformed the sampling problem into an equivalent optimization problem of finding a simpler distribution $q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)$ by minimizing ELBO.

If we impose no constraint on the approximation $q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)$, minimizing ELBO is numerically intractable. The intractability is largely due to the correlations among the parameters. To enable the computational tractability, we adapt the idea of *mean-field variational family* (Blei, Kucukelbir and McAuliffe (2017), Wand et al. (2011)), and impose an independence condition among the parameter components $\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2$ and $\sigma_{\mathbf{c}}^2$ as

(11)
$$q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2) = q(\mathbf{c}) q(\sigma_1^2) \cdots q(\sigma_{N_C}^2) q(\sigma_{\mathbf{c}}^2).$$

Under assumption (11), we can find an effective approximate posterior density, denoted by $q(\mathbf{c}, \sigma_1^2, \ldots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2)$ hereafter, using an alternating direction iterative algorithm, that is, at kth step, we solve the following optimization problems:

(12)
$$\begin{cases} \min_{q(\mathbf{c})} \mathrm{ELBO}(q(\mathbf{c}) q^{k-1}(\sigma_1^2) \cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2) | p(\bullet, \mathbf{R})), \\ \min_{q(\sigma_j^2)} \mathrm{ELBO}(q^k(\mathbf{c}) q^k(\sigma_1^2) \cdots q^k(\sigma_{j-1}^2) \\ \qquad \times q^{k-1}(\sigma_{j+1}^2) \cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2) | p(\bullet, \mathbf{R})), \quad j = 1, \ldots, N_C, \\ \min_{q(\sigma_{\mathbf{c}}^2)} \mathrm{ELBO}(q^k(\mathbf{c}) q^k(\sigma_1^2) \cdots q^k(\sigma_{N_C}^2) q(\sigma_{\mathbf{c}}^2) | p(\bullet, \mathbf{R})). \end{cases}$$

Here and later on, the (superscript and subscript) index k denotes the quantity of interest at kth iteration of the algorithm. Using condition (11), one can derive the explicit formulas for the minimizers of the optimization problems in (12) by examining the optimality system; that is, the following theorem holds.

THEOREM 1. *Assume that the conditional independence condition* (11) *holds. Then, the minimizers of the optimization problems in Algorithm* 1 *at each iteration have the following explicit formulas*:

(13)
$$
\begin{cases}
q^k(\mathbf{c}) = \mathcal{N}_+(\mathbf{c}_k, \Sigma_k^{\mathbf{c}}), \quad \mathbf{c}_k = \Sigma_k^{\mathbf{c}} \mathbf{K}^T \Sigma_k^{-1} \mathbf{R}, \\
\Sigma_k^{\mathbf{c}} = (\mathbf{K}^T \Sigma_k^{-1} \mathbf{K} + \sigma_{\mathbf{c},k}^{-2} \mathbf{L}^T \mathbf{L})^{-1}, \\
\Sigma_k^{-1} = \mathbb{E}_{q^{k-1}(\sigma_1^2) \cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)}[\Sigma^{-1}], \\
\sigma_{\mathbf{c},k}^{-2} = \mathbb{E}_{q^{k-1}(\sigma_1^2) \cdots q^{k-1}(\sigma_{N_C}^2) q^{k-1}(\sigma_{\mathbf{c}}^2)}[\sigma_{\mathbf{c}}^{-2}], \\
q^k(\sigma_j^2) = \mathrm{IG}\left(\sigma_j^2; \alpha_j + \dfrac{N_T}{2}, \beta_j + \dfrac{1}{2}\mathbb{E}_{q^k(\mathbf{c})}[(\mathbf{R}^j - \mathbf{K}^j \mathbf{c})^T(\mathbf{R}^j - \mathbf{K}^j \mathbf{c})]\right), \\
q^k(\sigma_{\mathbf{c}}^2) = \mathrm{IG}\left(\sigma_{\mathbf{c}}^2; \alpha_{\mathbf{c}} + \dfrac{n}{2}, \beta_{\mathbf{c}} + \dfrac{1}{2}\mathbb{E}_{q^k(\mathbf{c})}[(\mathbf{L}\mathbf{c})^T(\mathbf{L}\mathbf{c})]\right),
\end{cases}
$$

where the density function for truncated normal distribution $\mathcal{N}_+(\mathbf{c}^*, \Sigma_*^{\mathbf{c}})$ *is*

(14)
$$
\frac{1}{(1 - \mathrm{HN}(0))\sqrt{(2\pi)^n |\Sigma_*^{\mathbf{c}}|}} \exp\left\{-\frac{1}{2}(\mathbf{c} - \mathbf{c}^*)^T (\Sigma_*^{\mathbf{c}})^{-1}(\mathbf{c} - \mathbf{c}^*)\right\} \mathbf{1}_{\mathbf{c} \geq 0},
$$

$\mathrm{HN}(0) = \mathbb{P}(\xi \leq 0)$ *with* $\xi \sim \mathcal{N}(\mathbf{c}^*, \Sigma_*^{\mathbf{c}})$ *and* $\mathbb{E}_q[\bullet]$ *is the expectation with respect to the density* q.

In practice, the value of HN(0) can be estimated numerically. The proof of Theorem 1 follows a standard argument in theory of variational inference, and we provide a sketch in the Supplementary Material (Zhang et al. (2019)).

Now we discuss the convergence issue of scheme (12).

THEOREM 2. *The sequence* $\{q^k(\mathbf{c})q^k(\sigma_1^2) \cdots q^k(\sigma_{N_C}^2)q^k(\sigma_{\mathbf{c}}^2)\}$ *generated by scheme* (12) *converges, upon a subsequence, to a stationary point* $q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^*(\sigma_{\mathbf{c}}^2)$ *of the KL distance functional* $\mathcal{D}$, *which satisfies*

(15)
$$
\begin{cases}
q^*(\mathbf{c}) = \mathcal{N}_+(\mathbf{c}^*, \Sigma_*^{\mathbf{c}}), \\
q^*(\sigma_j^2) = \mathrm{IG}\left(\sigma_j^2; \alpha_j + \dfrac{N_T}{2}, \beta_j + \dfrac{1}{2}\mathbb{E}_{q^*(\mathbf{c})}[(\mathbf{R}^j - \mathbf{K}^j \mathbf{c})^T(\mathbf{R}^j - \mathbf{K}^j \mathbf{c})]\right), \\
q^*(\sigma_{\mathbf{c}}^2) = \mathrm{IG}\left(\sigma_{\mathbf{c}}^2; \alpha_{\mathbf{c}} + \dfrac{n}{2}, \beta_{\mathbf{c}} + \dfrac{1}{2}\mathbb{E}_{q^*(\mathbf{c})}[(\mathbf{L}\mathbf{c})^T(\mathbf{L}\mathbf{c})]\right), \\
\mathbf{c}^* = \Sigma_*^{\mathbf{c}} \mathbf{K}^T \Sigma_*^{-1} \mathbf{R}, \quad \Sigma_*^{\mathbf{c}} = (\mathbf{K}^T \Sigma_*^{-1} \mathbf{K} + \sigma_{\mathbf{c},*}^{-2} \mathbf{L}^T \mathbf{L})^{-1}, \\
\Sigma_*^{-1} = \mathbb{E}_{q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2) q^*(\sigma_{\mathbf{c}}^2)}[\Sigma^{-1}], \\
\sigma_{\mathbf{c},*}^{-2} = \mathbb{E}_{q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2) q^*(\sigma_{\mathbf{c}}^2)}[\sigma_{\mathbf{c}}^{-2}].
\end{cases}
$$

We refer readers to Appendix B for the proof of Theorem 2. To that end, let us consider the stopping principle of our scheme. Various stopping criteria exist for an iterative algorithm, for example, ELBO is usually calculated (whenever possible)

Algorithm 1 A variational Bayesian algorithm for solving (5)

Input: Model parameters $\{\alpha_j, \beta_j\}_{j=1}^{N_C}$ and $\alpha_{\mathbf{c}}$, $\beta_{\mathbf{c}}$. Initial guesses of $q^0(\mathbf{c})$. Tolerance ε.

Output: The approximate posterior distribution of $p(\mathbf{c})$ is $q^k(\mathbf{c})$.

1: $k \leftarrow 1$
2: **while** $\Delta_i(q^k(\mathbf{c})) > \varepsilon$ **do**
3:     Update the distributions of hyperparameters $\sigma_{j,k}^2$ and $\sigma_{\mathbf{c},k}^2$ by formula (15), replacing $\mathbf{c}$ by $\mathbf{c}_k$.
4:     Update the distribution of solution $\mathbf{c}_k$ by formula (13).
5:     $k \leftarrow k+1$
6: **end while**

---

as a stopping criterion in the variational inference. For our problem, a natural and simple stopping criterion is to utilize the mean and covariance of the interested quantity—the rate constant map $\mathbf{c}$. Hence, we define the accuracy of our algorithm as

$$(16) \qquad \Delta_1(q^k(\mathbf{c})) = \frac{\|\mathbf{c}_k - \mathbf{c}_{k-1}\|}{\|\mathbf{c}_{k-1}\|}, \qquad \Delta_2(q^k(\mathbf{c})) = \frac{\|\Sigma_{*,k}^{\mathbf{c}} - \Sigma_{*,k-1}^{\mathbf{c}}\|}{\|\Sigma_{*,k-1}^{\mathbf{c}}\|}.$$

Finally, by coupling the stopping criteria and Theorem 1, we offer a variational Bayesian algorithm for reconstructing the rate constant map in the biosensor system in Algorithm 1.

3.3. *An adaptive strategy to improve the quality of the rate constant map.* As mentioned in Remark 1, we start with a low dimensionality n of solution space where the rate constant map f is only estimated on a set of coarse distributed points in Ω. Therefore, the resolution of the rate constant map is quite low. In order to improve the quality of the estimated rate constant map f_n, we adopt the idea of the oriented adaptive discretization, from numerical technique in finite element methods (Chen, Holst and Xu (2009), Zhang et al. (2018)). The idea, combined with our approach, is to first solve the problem (1) through the variational Bayesian approach (Algorithm 1), obtaining the solution (the distribution of the random vector $\mathbf{c}$) to the current triangulation. The quality is then estimated using the solution, and is used to mark a set of triangles to be refined. Triangles are refined in a way that maintains two of the most important properties of the triangulations: shape regularity and conformity.

Given an estimated distribution $\mathcal{N}_+(\mathbf{c}^*, \Sigma_*^{\mathbf{c}})$, let $\widehat{\mathbf{c}}$, $\underline{\mathbf{c}}$ and $\overline{\mathbf{c}}$ be the sample mean, the lower and upper endpoints of the 95% confidence interval by samples from Algorithm 1. Define by $V_\mu^{\text{mean}}[\widehat{\mathbf{c}}]$ the variation of the mean vector $\widehat{\mathbf{c}}$ over the elements $\triangle_\mu$, that is, $V_\mu^{\text{mean}}[\widehat{\mathbf{c}}] = \frac{1}{3} \sum_{\iota \neq J, \iota, J \in \{1,2,3\}} |\widehat{\mathbf{c}}_{\mu_\iota} - \widehat{\mathbf{c}}_{\mu_J}|$. Similarly, we can define the

---

Algorithm 2 An adaptive variational Bayesian approach (AVBA) for reconstructing the rate constant map in the biosensor system

---

Input: Initial mesh: $\mathcal{T}_1$. Tolerance number: τ. Tolerance error: ε. Sample number: N. The maximum iteration number $K_{\max}$.

Output: The estimated rate constant mean map: $\widehat{\mathbf{c}} = \widehat{\mathbf{c}}_k$. The lower and upper rate constant maps: $\underline{\mathbf{c}} = \underline{\mathbf{c}}_k$ and $\overline{\mathbf{c}} = \overline{\mathbf{c}}_k$.

1: $k \leftarrow 1$
2: **while** $\Delta_i(q^k(\mathbf{c})) > \varepsilon$ **and** $k < K_{\max}$ **do**
3:     Obtain the distribution $\mathcal{N}_+(\mathbf{c}_k, \Sigma_k^{\mathbf{c}})$ of random vector $\mathbf{c}$ of the finite element solution of problem (1) on the mesh $\mathcal{T}_k$ by Algorithm 1.
4:     Compute the sample mean $\widehat{\mathbf{c}}_k$, the lower confidence limit $\underline{\mathbf{c}}_k$ and the upper confidence limit $\overline{\mathbf{c}}_k$ using samples for $\mathcal{N}_+(\mathbf{c}_k, \Sigma_k^{\mathbf{c}})$ from Algorithm 1.
5:     Refine the mesh $\mathcal{T}_k$ at all points where $V_\mu \geq \tau \max_{\triangle_\mu \in \mathcal{T}} V_\mu$.
6:     Construct a new mesh $\mathcal{T}_{k+1}$.
7:     $k \leftarrow k + 1$
8: **end while**

---

variation of $\underline{\mathbf{c}}$ and $\overline{\mathbf{c}}$, denoted by $V_\mu^{\text{lower}}[\underline{\mathbf{c}}]$ and $V_\mu^{\text{upper}}[\overline{\mathbf{c}}]$ respectively. Finally, define the refinement indicator V_μ by

$$(17) \qquad V_\mu[\widehat{\mathbf{c}}, \underline{\mathbf{c}}, \overline{\mathbf{c}}] = \max\{V_\mu^{\text{mean}}[\widehat{\mathbf{c}}], V_\mu^{\text{lower}}[\underline{\mathbf{c}}], V_\mu^{\text{upper}}[\overline{\mathbf{c}}]\}.$$

Then, the refinement should be done in the triangles of all points in the finite element mesh $\mathcal{T}$ where the function V_μ achieves its maximum; that is, refine the mesh in such triangles of Ω where

$$(18) \qquad V_\mu[\widehat{\mathbf{c}}, \underline{\mathbf{c}}, \overline{\mathbf{c}}] \geq \tau \max_{\triangle_\mu \in \mathcal{T}} V_\mu[\widehat{\mathbf{c}}, \underline{\mathbf{c}}, \overline{\mathbf{c}}],$$

where $\tau \in (0, 1)$ are numbers which should be chosen computationally.

Finally, we use the longest edge refinement rule for dividing the marked triangles, so that the mesh obtained by this dividing rule still conforms and has a shape regular (Rivara (1984)). Note that apart from the marked triangles, the additional triangles are also refined to recover the conformity of triangulations. In this work, we also control the number of elements added to ensure the overall optimality of the refinement procedure.

Let us present our main algorithm for solving the two-dimensional Fredholm integral equation of the type (1) in Algorithm 2.

4. Simulation study. This section presents the numerical results for the inverse problem (1), of which the adaptive variational Bayesian approach (AVBA), that is, Algorithm 2, is used to illustrate its accuracy and flexibility. The simulation consists of three steps. First, a simulated response function R_{obs} is generated by computer according to (1) for a given rate constant map $\bar{f}(k_a, k_d) =$
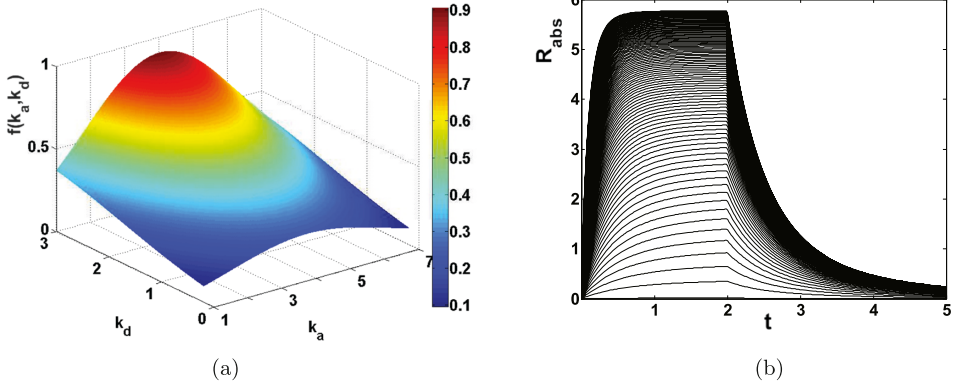
FIG. 4. (a) *The exact rate constant map for model problem.* (b) *The synthetic time dependent noise-free data.*

$\exp(-0.1 * ((k_a - 4)^2 + (k_d - 4)^2))$ in the domain $\Omega = [1, 7] \times [0, 3]$—see (a) in Figure 4. Denote by $\{\bar{R}_{p,q}\}_{p,q=1}^{N_T, N_C}$ (see (b) in Figure 4) the collected exact data at a uniform time grid $\{t_p\}_{p=1}^{N_T}$ ($N_T = 300$, $T = [0, 4]$, $t_0 = 0$ and $t_{inj} = 2$), at a uniform analyte concentration grid $\{C_q\}_{q=1}^{N_C}$ ($N_C = 100$ and $\Xi = [0.001, 2]$). The synthetic noisy data is generated by

$$R_{\mathrm{obs}, p, q}^{\delta} = \bar{R}_{p,q} + \max\left\{0, \max_p \bar{R}_{p,q} \delta \omega_q\right\}, \quad p = 1, \dots, N_T, q = 1, \dots, N_C,$$

where δ indicates the noise level, and ω_q is a Gaussian random variable with zero mean and standard deviation $\sqrt{q}$. At the last step of the simulation, the observation data $R_{\mathrm{obs}}^{\delta}$ is processed through our algorithm, and the retrieved rate constant map $f(k_a, k_d)$ is compared with the input map. To assess the accuracy of the approximate solution, we define the L_2-norm relative error for an estimated random vector $\mathbf{c}$ with distribution $\mathcal{N}_+(\hat{\mathbf{c}}^*, \Sigma_*^{\mathbf{c}})$ as

$$L2Err = \frac{\|\bar{f}(k_a, k_d) - \hat{f}(\hat{\mathbf{c}}^*)\|_{L^2(\Omega)}}{\|\bar{f}(k_a, k_d)\|_{L^2(\Omega)}}, \quad \hat{f}(\hat{\mathbf{c}}^*) = \sum_{l=1}^{n} \hat{\mathbf{c}}_l^* \cdot \varphi_l(k_a, k_d).$$

A demonstration of Algorithm 2 can be found in the Supplementary Material. The fundamental assumption of our approach, according to (11), is the conditional independence of random variables $\{\mathbf{c}, \sigma_1^2, \dots, \sigma_{N_C}^2, \sigma_{\mathbf{c}}^2\}$ given the data $\mathbf{R}$. To examine the assumption (11), we compute the Pearson correlation coefficients $\rho_{\mathbf{c}, \sigma_j^2}$ ($j = 1, \dots, N_C$), $\rho_{\mathbf{c}, \sigma_{\mathbf{c}}^2}$ and $\rho_{\sigma_i^2, \sigma_{\mathbf{c}}^2}$ ($j = 1, \dots, N_C$) between the vectors $\mathbf{c}$, $\sigma_{\mathbf{c}}^2$ and σ_j^2 respectively, from the MCMC samples. The correlation coefficient between the vector $\mathbf{c}$ and scalars are computed componentwise, that is, with $\mathbf{c}_i$. The results, for the example in the Supplementary Material, are shown in Figure 5, where the abscissa i (or j) denotes the ith (or jth) component. Overall, the correlation coefficients between $\sigma_{\mathbf{c}}^2$ and σ_i^2 are very small, with a maximum norm

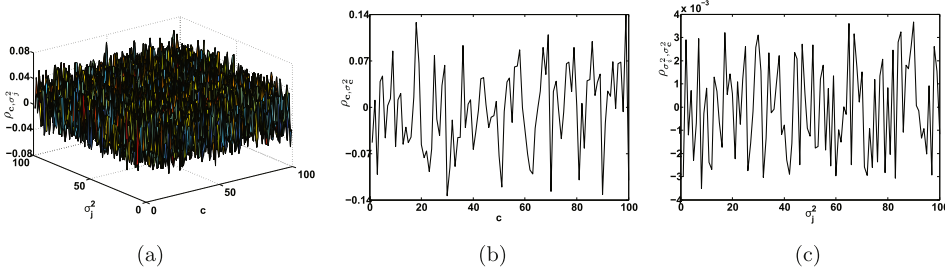(a)　　　　　　　　　　　　(b)　　　　　　　　　　　　(c)

FIG. 5. *Correlation coefficients $\rho_{\mathbf{c},\sigma_j^2}$ $(j = 1, \ldots, N_C)$, $\rho_{\mathbf{c},\sigma_{\mathbf{c}}^2}$ and $\rho_{\sigma_i^2,\sigma_{\mathbf{c}}^2}$ $(j = 1, \ldots, N_C)$ between the vectors $\mathbf{c}$, $\sigma_{\mathbf{c}}^2$ and σ_j^2. The noise level $\delta = 0.001$. (a) Correlation coefficients $\rho_{\mathbf{c},\sigma_j^2}$. (b) Correlation coefficients $\rho_{\mathbf{c},\sigma_{\mathbf{c}}^2}$. (c) Correlation coefficients $\rho_{\sigma_i^2,\sigma_{\mathbf{c}}^2}$.*

$|\rho_{\sigma_i^2,\sigma_{\mathbf{c}}^2}|_\infty := \max_{1 \le i \le N_C} |\rho_{\sigma_i^2,\sigma_{\mathbf{c}}^2}|$, smaller than 0.004 for all three noise levels. The correlation coefficients between $\mathbf{c}$ and σ_i^2 are also small with a maximum norm $|\rho_{\mathbf{c},\sigma_j^2}|_\infty := \max_{1 \le i \le n, 1 \le j \le N_C} |\rho_{\mathbf{c}_i,\sigma_j^2}|$ smaller than 0.08 for all three noise levels. The correlation coefficients between $\mathbf{c}$ and $\sigma_{\mathbf{c}}^2$ are slightly larger with the maximum norm $|\rho_{\mathbf{c},\sigma_{\mathbf{c}}^2}|_\infty := \max_{1 \le i \le n} |\rho_{\mathbf{c}_i,\sigma_{\mathbf{c}}^2}|$ close to 0.14. Hence, we can conclude that the correlation between $\mathbf{c}$, $\sigma_{\mathbf{c}}^2$ and σ_j^2 is relatively weak.

Now we investigate the robustness of Algorithm 2. We first study the sensitivity of the numerical results with respect to the prior parameters $\{\alpha_j, \beta_j\}_{j=1}^{N_C}$ and $\{\alpha_{\mathbf{c}}, \beta_{\mathbf{c}}\}$, and illustrate the rationale behind these choices in the following examples. If the value of α_j, $\alpha_{\mathbf{c}}$ is not too large, the distributions of σ_j^2 and $\sigma_{\mathbf{c}}^2$ in (15) are dominated by $N_T/2$ and $n/2$ in the numerators $\alpha_j + N_T/2$ and $\alpha_{\mathbf{c}} + n/2$, and can thus only have a marginal impact on the algorithm. Therefore, we only study the influence of parameters β_j, $\beta_{\mathbf{c}}$. Since $\{\beta_j\}_{j=1}^{N_C}$ plays the same role in the algorithm, we only present the result for β_1 here. The results are displayed in Table 1,

TABLE 1
Sensitivity analysis for parameters $\beta_{\mathbf{c}}$ and β_1. $\delta = 1\%$. $\beta_j = 1$ for $j \ne 1$. $\alpha_j = \alpha_{\mathbf{c}} = 1$

$\beta_{\mathbf{c}}$	β_1	L2Err	$|\rho_{\sigma_i^2,\sigma_{\mathbf{c}}^2}|_\infty$	$|\rho_{\mathbf{c},\sigma_j^2}|_\infty$	$|\rho_{\mathbf{c},\sigma_{\mathbf{c}}^2}|_\infty$
0.0001	0.0001	0.0661	0.0032	0.0130	0.2247
0.01	0.01	0.0659	0.0002	0.0304	0.0414
1	1	0.0658	0.0029	0.0232	0.2175
100	100	0.0661	0.0021	0.0258	0.1502
10,000	10,000	0.0660	0.0038	0.0268	0.0931
0.0001	10,000	0.0664	0.0092	0.0268	0.1406
0.01	100	0.0667	0.0020	0.0836	0.2609
0.1	10	0.0660	0.0021	0.0135	0.3592

TABLE 2
Sensitivity analysis for the initial guess of $q^0(\mathbf{c})$ with fixed type of distribution
$\mathcal{N}((\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}, \kappa\mathbf{I})$. $\delta = 1\%$. $\alpha_j = \beta_j = \alpha_{\mathbf{c}} = \beta_{\mathbf{c}} = 1$

λ	κ	L2Err	$\|\rho_{\sigma_i^2, \sigma_{\mathbf{c}}^2}\|_\infty$	$\|\rho_{\mathbf{c}, \sigma_j^2}\|_\infty$	$\|\rho_{\mathbf{c}, \sigma_{\mathbf{c}}^2}\|_\infty$
0.00001	0.0001	0.0078	0.0008	0.0281	0.0162
0.00001	0.01	0.0080	0.0026	0.0143	0.1272
0.00001	1	0.0125	0.0016	0.0137	0.0381
0.00001	10	0.0531	0.0022	0.0101	0.2360
0.001	0.01	0.0517	0.0009	0.0061	0.2440
0.1	0.01	0.0621	0.0019	0.0145	0.0828
1	0.01	0.2179	0.0098	0.0178	0.1738
100	0.01	0.4512	0.0203	0.0782	0.3491

where the data noise level equals 1%, and parameters β_j ($j > 1$), $\{\alpha_j\}$ and $\alpha_{\mathbf{c}}$ are all set as 1. The results show that both the estimated rate constant mean map and the conditional independence assumption (11) seem relatively independent of the parameters $\beta_{\mathbf{c}}$ and β_1 for the magnitude of the relative error L2Err, and the correlation coefficients $\|\rho_{\sigma_i^2, \sigma_{\mathbf{c}}^2}\|_\infty$, $\|\rho_{\mathbf{c}, \sigma_j^2}\|_\infty$ and $\|\rho_{\mathbf{c}, \sigma_{\mathbf{c}}^2}\|_\infty$ remain almost unchanged as the value of $\beta_{\mathbf{c}}$ and β_1 varies.

We next consider the influence of the initial guess of the distribution for the rate constant map $q^0(\mathbf{c})$. In Table 2 we investigate the case with the Gaussian prior. The expectation is assumed as the Tikhonov regularized solution $\mathbf{c}^* = (\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}$, where the regularization parameter λ varies from 0.00001 to 100. The covariance matrix is assumed to be of the type $\kappa\mathbf{I}$. The first four rows in Table 2 show that with a small value of λ the relative error is small for $\kappa \leq 1$. The last four rows in Table 2 show that the smaller the value of λ, the better the estimated rate constant mean map. Other types of mean $\mathbf{c}_0$ and covariance matrix $\Sigma_{\mathbf{c}}$ were also tested for this example. Worse results are obtained for other types of mean $\mathbf{c}_0$. However, arbitrary types of covariance matrix $\Sigma_{\mathbf{c}}$ with small values of $\|\Sigma_{\mathbf{c}}\| \leq 1$ give the same good results if one chooses the mean $\mathbf{c}_0$ appropriately. Therefore, based on this group of simulations, we suggest using the initial guess of $q^0(\mathbf{c})$ as a Gaussian distribution with mean $(\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}$ ($\lambda \leq 0.001$) and covariance matrix $\Sigma_{\mathbf{c}}$ ($\|\Sigma_{\mathbf{c}}\| \leq 1$).

Now we investigate the stability of the estimated solutions with respect to the noise level δ. Specifically, set $\delta = 0.001, 0.005, 0.01, 0.02$ and implement Algorithm 2 repeatedly. The initial guess of expectation is set as the Tikhonov regularized solution $\mathbf{c}^* = (\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}$, where the regularization parameter λ is chosen by the L-curve method Engl, Hanke and Neubauer (1996). The pairs of parameters $\{(\alpha_j, \beta_j)\}_{j=1}^{N_C}$ and $(\alpha_{\mathbf{c}}, \beta_{\mathbf{c}})$ for the inverse Gamma distribution are taken to be $(1, 1)$. The relative errors of the estimate rate constant mean map and

TABLE 3

Sensitivity analysis of the algorithm with different noisy data $\mathbf{c}_0 \sim \mathcal{N}((\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}, \mathbf{I})$.
$\alpha_j = \beta_j = \alpha_{\mathbf{c}} = \beta_{\mathbf{c}} = 1$

δ	L2Err	$|\rho_{\sigma_i^2, \sigma_{\mathbf{c}}^2}|\infty$	$|\rho_{\mathbf{c}, \sigma_j^2}|\infty$	$|\rho_{\mathbf{c}, \sigma_{\mathbf{c}}^2}|\infty$
0.001	0.0098	0.0011	0.0263	0.0459
0.005	0.0111	0.0024	0.0193	0.3285
0.01	0.0318	0.0023	0.0540	0.1941
0.02	0.1520	0.0015	0.0164	0.5070
0.10	0.4365	0.0037	0.0313	0.1497
0.20	0.4253	0.0029	0.0462	0.2979
0.40	0.5049	0.0018	0.0440	0.3387

the correlation between parameters are displayed in Table 3. The estimated lower and upper rate constant maps for different noise data are shown in Figure 6. These results demonstrate that our algorithm is stable with respect to a small noise error ($\delta \leq 0.02$). It should be pointed out that with a large noise data, based on $\delta \geq 0.1$, the estimated rate constant mean map is quite poor and the value of $\|\Sigma_*^{\mathbf{c}}\|$ is relatively large, though the algorithm is still convergent with the given system parameters.

Finally, we compare our approach with the MCMC. For simplicity, we fix the grid with 450 triangles and 256 nodes, and employ Algorithm 2 and the MCMC with the same initial guess of parameters ($\alpha_j = \beta_j = \alpha_{\mathbf{c}} = \beta_{\mathbf{c}} = 1$ and $\mathbf{c}_0 \sim \mathcal{N}((\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}, \mathbf{I})$). The resulting posterior distribution (8) is sampled using the standard Gibbs sampler, and the length of the Markov chain is 50,000, with the first 5000 realizations discarded as transient states. The mixing of the Markov chain is monitored by visually inspecting the trace plot and calculating the correlation coefficient. Note that we set a maximal implement time of $T_{\max} = 300$ min. in all of our simulations. We obtained an output of the result even in the occurrence of the maximal time point of the algorithm. Numerical results with different noise levels are shown in Figure 7 and Table 4. Furthermore, in order to quantitatively evaluate the difference between our method and MCMC, we compute their relative approximate Wasserstein distance W, which is defined as follows: let Ω_j^{VB} and Ω_j^{MCMC} be the samples of the two groups (sample sizes are different) of $\mathbf{c}_j$ from our method and MCMC, respectively. Then, the relative approximate Wasserstein distance for $\mathbf{c}_j$ between our method and MCMC is defined as

$$W_j := \frac{1}{100} \sum_{i=1}^{100} |\tilde{c}_j^{\mathrm{VB},i} - \tilde{c}_j^{\mathrm{MCMC},i}| / (\tilde{c}_j^{\mathrm{MCMC},i} + 1),$$

where $\tilde{c}_j^{\mathrm{VB},i}$ and $\tilde{c}_j^{\mathrm{MCMC},i}$ represent the $i\%$ quantiles from two sample sets Ω_j^{VB} and Ω_j^{MCMC}, respectively. The calculated relative approximate Wasserstein distances
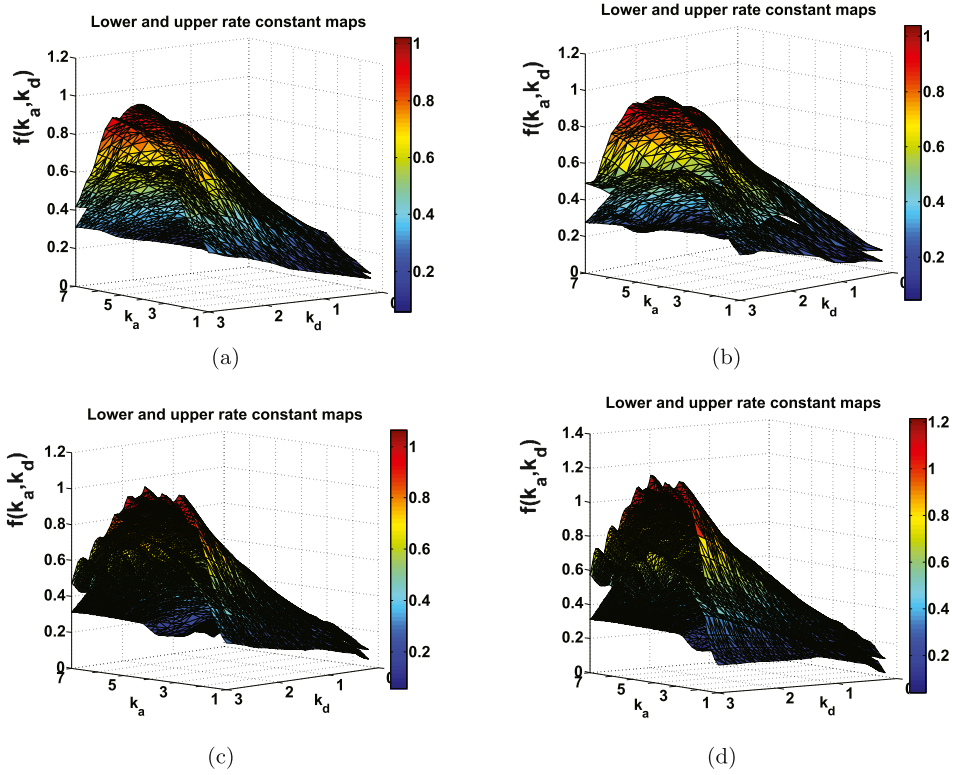
FIG. 6. *The estimated lower and upper rate constant maps for different noise data. The initial triangulation*: *the number of nodes equals* 100, *and the number of triangles is* 162. (*a*) $\delta = 0.001$. *The final triangulation*: *the number of nodes equals* 1261, *and the number of triangles is* 2396. (*b*) $\delta = 0.005$. *The final triangulation*: *the number of nodes equals* 1146, *and the number of triangles is* 2166. (*c*) $\delta = 0.01$. *The final triangulation*: *the number of nodes equals* 3930, *and the number of triangles is* 7652. (*d*) $\delta = 0.02$. *The final triangulation*: *the number of nodes equals* 4566, *and the number of triangles is* 8914.

for all $\mathbf{c}_j$ ($j = 1, \ldots, 256$) with different data sets are presented in Figure 8(a). The histograms of $\mathbf{c}_{100}$ by the data set with $\delta = 0.01$ for both our method and MCMC are provided in Figure 8(b). All of these outputs indicate that the results obtained using our method and the MCMC almost coincided for the small noise level, while for large noisy data the estimated mean map achieved through our method was much better than the results acquired using the MCMC. It may even be possible (e.g., in a case with noise level $\delta = 0.4$ in Table 4) that the MCMC does not converge, while our algorithm converges quickly. Moreover, as one can see in Table 4, it clearly shows the computational efficiency of our method over the MCMC for all situations. Finally, it should be noted that when the problem size is large, for example, $n > 1000$, for the majority of cases the MCMC cannot offer a satisfactory result within a reasonable time (e.g., two days). Instead, our
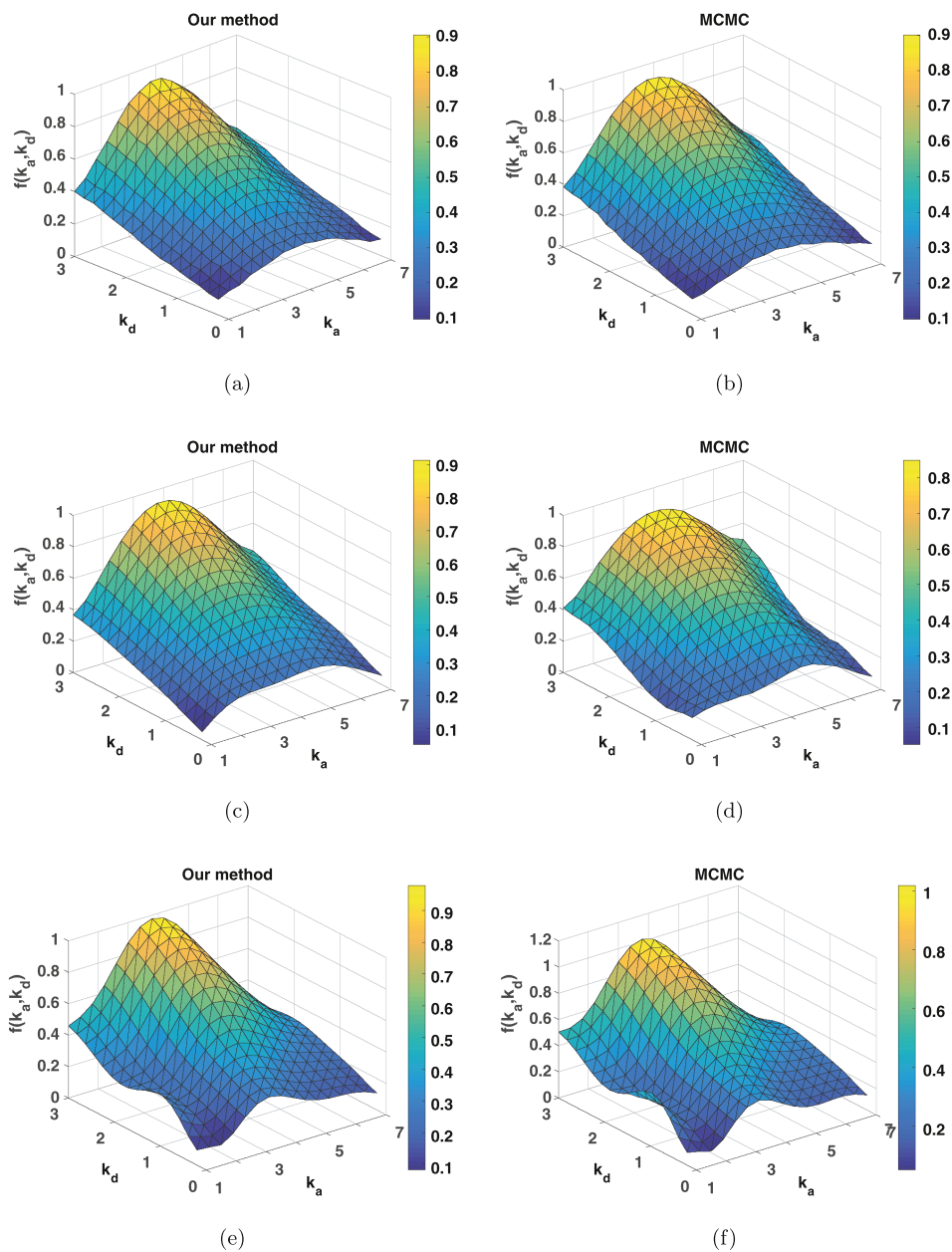
FIG. 7.   *A comparison of our method (Algorithm* 2*) with MCMC. (a) Algorithm* 2 *for* δ = 0.002. *(b) MCMC for* δ = 0.002. *(c) Algorithm* 2 *for* δ = 0.01. *(d) MCMC for* δ = 0.002. *(e) Algorithm* 2 *for* δ = 0.01. *(f) MCMC for* δ = 0.01. *(g) Algorithm* 2 *for* δ = 0.01. δ = 0.05. *(h) MCMC for* δ = 0.05.

TABLE 4
Relative error and running time of the estimated rate constant mean map for our method and MCMC. $\mathbf{c}_0 \sim \mathcal{N}((\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{K}^T\mathbf{R}, \mathbf{I})$. $\alpha_j = \beta_j = \alpha_{\mathbf{c}} = \beta_{\mathbf{c}} = 1$

	Relative error		Running time (min.)	
δ	MCMC	AVBA	MCMC	AVBA
0.002	0.0024	0.0022	21.18	1.75
0.01	0.0112	0.0098	29.45	3.01
0.05	0.0638	0.0232	25.77	3.68
0.10	0.2840	0.1632	73.12	7.23
0.20	0.5114	0.2915	49.94	5.68
0.40	3.8116	0.7451	$T_{\max}$	7.68

method provided a convergent result within a short time. For instance, in our real data application, when $n = 1812$, the implementation of our method took only 42 minutes.

5. A real data application. In this section, our method (AVBA) is tested on real experimental data—parathyroid hormone (PTH). In the experiment, the human PTH1R receptor was immobilized on a LNB-carboxyl biosensor chip using amine coupling according to the manufacturer's instructions. Using the flow rate 25 μL/min at 20.0°C, we did 35 μL injections of the peptide PTH(1-34) at six concentration levels from 1214 nM to 9714 nM; see the solid lines in (a) of Figure 9. The sensorgrams were measured using a QCM biosensor Attana Cell 200 (Attana AB, Stockholm, Sweden) instrument.
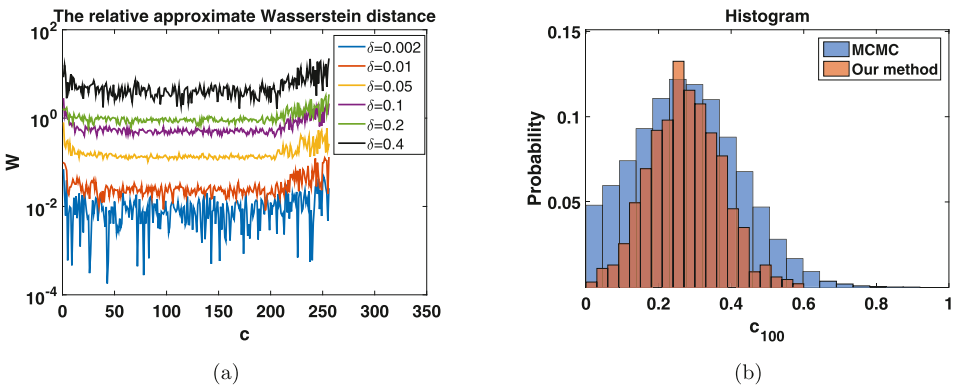

FIG. 8. (*a*) *The relative approximate Wasserstein distance between our method the MCMC of* $\mathbf{c}$ *for data sets with different noise level* δ. (*b*) *Histograms for* $\mathbf{c}_{100}$ *by the data set with* $\delta = 0.01$ *for our method* (2735 *samples*) *and MCMC* (50,000 *samples*). *The difference between two histograms in* (*b*) *corresponds to a point with the abscissa* 100 *on the red curve in* (*a*).
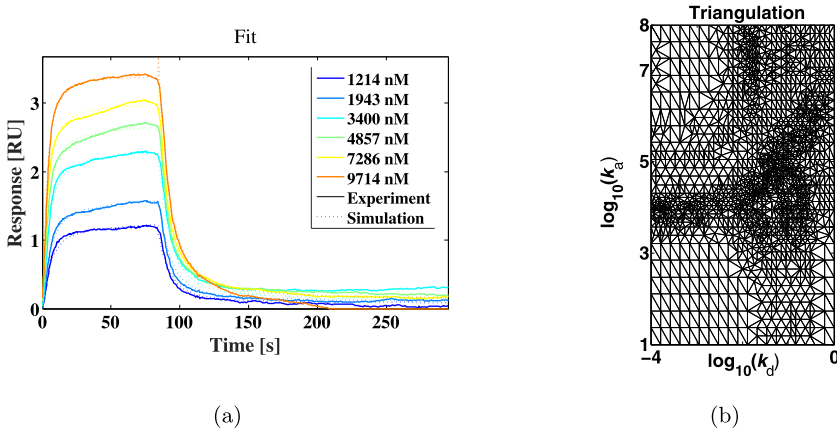
(a)

(b)

FIG. 9. (*a*) *The experimental data* (*solid line*) *and simulated response* (*dashed line*) *with different concentrations of PTH at the temperature of* 20 °*C*. (*b*) *Triangulations at the final iteration. The number of nodes equals* 1812, *and the number of triangles is* 3498.

The initial triangulation is uniformly distributed in the log-scale domain $(\log_{10}(k_d), \log_{10}(k_a)) \in [-4, 0] \times [1, 8]$ with $20 \times 20 = 400$ node points. Algorithm 2 stopped at the 9th iteration with 1812 nodes and 3498 triangles—see (b) in Figure 9. The estimated rate constant mean map and corresponding intensity map are shown in Figure 10. In (a) of Figure 9, we show a comparison between the experimental data (solid line) and simulated response curves (dashed line), which is obtained by solving the forward problem with the estimated rate constant mean solution. The results with approximately 96% overlap show that the estimated rate constant mean map can be used as the real rate constant map in a deterministic model.
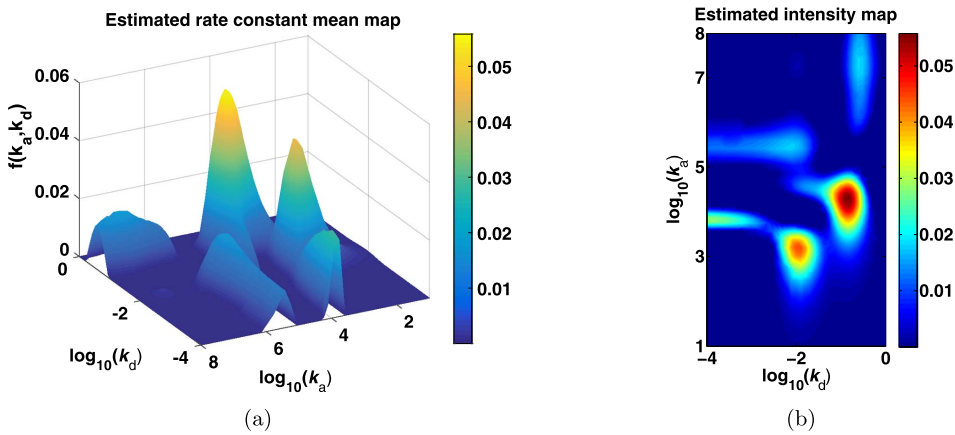

(a)

(b)

FIG. 10. (*a*) *The estimated rate constant mean map*. (*b*) *The intensity map*.

As discussed in the Introduction, the peaks form the most important piece of information in the rate constant map, as they enable the conclusion of the number of interactions as well as the value of the active association and dissociation constants for the given large molecular system. These peaks can be considered as the geometry representation of the Intrinsic Property of the Interaction (IPoI). Note that the rate constant mean map in Figure 10, as well as the similar maps in Svitel et al. (2003), Altschuh et al. (2012), and Gorshkova et al. (2008), contains too much redundant information about a chemical reaction besides the IPoI. Though we can derive the reaction information from the mean maps (see Figure 10) via ad hoc work—as demonstrated in Svitel et al. (2003), Altschuh et al. (2012), Gorshkova et al. (2008) and Zhang et al. (2018)—to improve the accuracy and efficiency of the estimation, we take an alternative perspective in this work; that is, we provide two new approaches that automatically derive the IPoI of the biosensor system.

The first approach is a thresholding method for the mean map, which will be named the ν-Thresholding Contour Method (TCM). This approach cuts off all regions where the value of the rate constant mean map is less than the $1 - \nu\%$ of the maximum value of the estimated rate constant. In Figure 11, we display the results of TCM with different thresholds ν. Actually, with $\nu \in [2, 8]$, the TCM provides exactly two isolated regions with some contours. We point out that the value of ν might vary case by case for the different datasets. However, $\nu \in [4, 6]$ seems to work well for all tested data. Therefore, we recommend using $\nu = 5$ in practice. From (c) of Figure 11, we conclude that two reactions exist for our PTH system, and their positions are $(\log_{10}(k_d), \log_{10}(k_a)) = (-2.0, 3.2)$ and $(-0.9, 4.4)$.

Now consider our second approach to automatically determine the IPoI. In this approach, instead of using the mean map, we consider the other moment maps. Figure 12 displays the different moment maps from the 0.5th moment map to the fourth moment map. As can be seen in the evolution of the moment maps in Figure 12, the higher the moment, the sharper the map. For the map higher than the

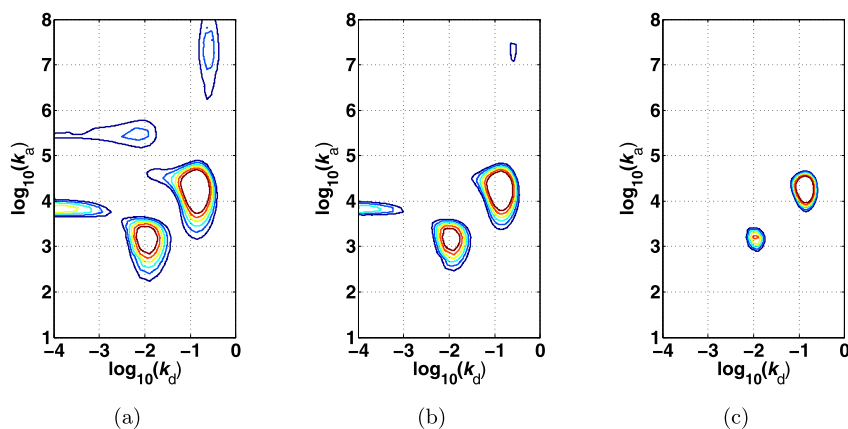FIG. 11.   *TCM with different thresholds ν: $\nu = 15$ in (a), $\nu = 10$ in (b), $\nu = 5$ in (c).*
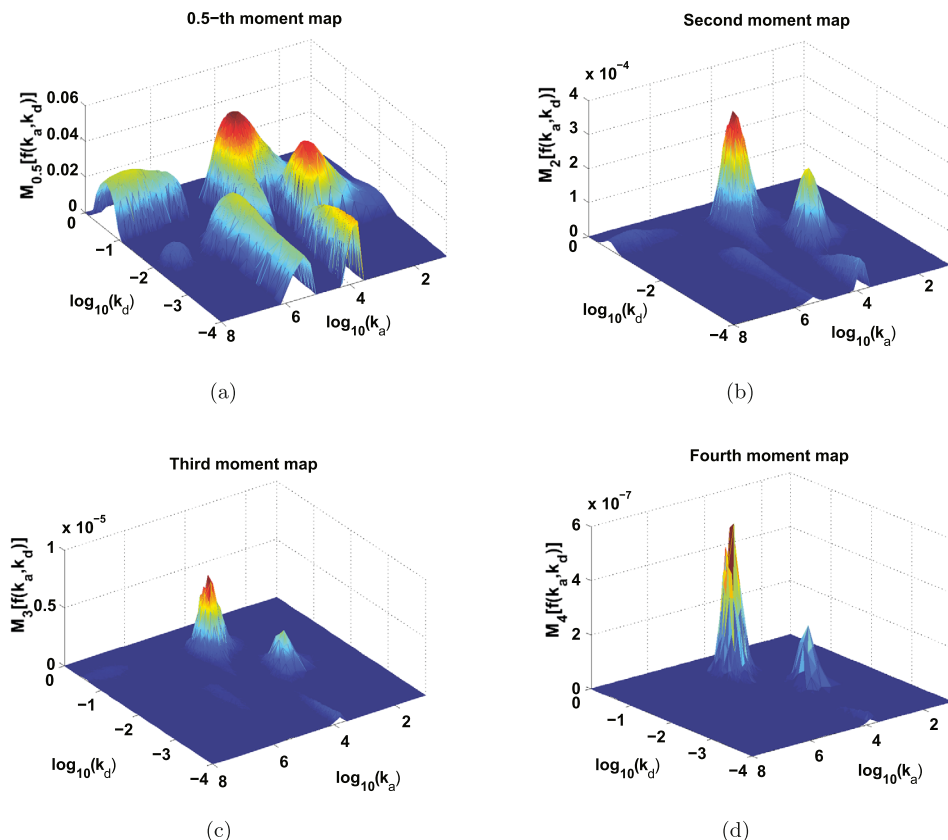
FIG. 12. *Moment maps*: (*a*) 0.5*th moment map*. (*b*) *Second moment map*. (*c*) *Third moment map*. (*d*) *Fourth moment map*.

third moment, the moment map only has two peaks, and the other flatter local peaks that appear in the lower moment maps vanish. This phenomenon implies that the high moment map filters the redundant information, and highlights the interaction information. From this, we can conclude that the high moment map represents IPoR in a more efficient way. However, the third moment map provides almost the full picture of the IPoR, that is, moment maps beyond that might not be necessary; see Figure 13, which shows the similarity between the third and fourth moment intensity maps. The high order moment intensity maps—see Figure 13 ((a) for the third moment and (b) for the fourth moment)—indicate that there are two interactions in the PTH system. The corresponding association and dissociation rate constants are spotted at $(-2.0, 3.2)$ and $(-0.9, 4.4)$ respectively. These results coincide with the conclusion obtained by the TCM.

Finally, we remark that in this real data application, we did not show the lower and upper rate constant maps since they present very similar sharps to the mean map. Hence, the lower and upper rate constant maps do not contain more infor-
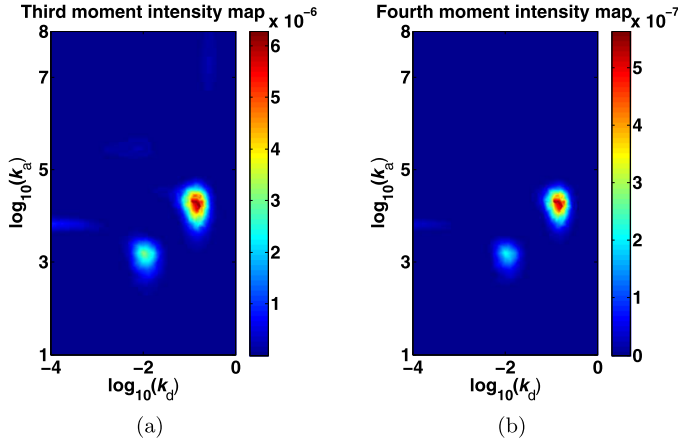
FIG. 13. *High order moment intensity maps*: (*a*) *Third moment intensity map.* (*b*) *Fourth moment intensity map.*

mation about the IPoR of the PTH system. However, for other biosensor systems, we remain inclined to suggest performing this analysis since it is possible that the sharps of the lower and upper rate constant maps behave differently, which will be interesting to study when looking at the advanced properties of IPoR for biosensor systems.

6. Conclusion. In this paper, an inverse problem of estimating the interaction number as well as their rate constants in biosensor systems has been studied in detail. Unlike the conventional deterministic mathematical models in the field of biosensor systems, we propose a statistical model based on the Bayesian framework. By engaging a large number of hyper parameters in the model, the full Bayesian solution results in a very high dimensional parameter space, which leads to a serious computational problem in terms of sampling the posterior distribution of the rate constants. To overcome this obstacle, we develop an Adaptive Variational Bayesian Approach (AVBA) for estimating the rate constants in biosensor systems. Numerical examples for both synthetic and real data demonstrate that the AVBA is stable and efficient. By combining the approach with the thresholding contour method or high order moment map method, the AVBA has shown its potential in capturing the intrinsic property of the interaction of a biosensor system; that is, it can provide us with a method to automatically find the number of interactions in a biosensor system, and the value of association and dissociation rate constants corresponding to these interactions.

It should be noted that the AVBA is a single step method that could accurately resolve the two underlying interactions without a priori assumptions of the existence of a parallel interactions or the range of expected kinetic parameters, while

the classical regression analysis (the parallel reactions model) of the kinetic sets produces correct results depending on the starting dissociation (e.g., k_d in this paper) values. Although more experience with other biosensor systems would be needed to better understand the potential and limitations of the AVBA, its applications in several artificial problems and the PTH system seem very promising. By exploiting the full kinetic data set available, the obtained two-dimensional kinetic and affinity distributions have a higher resolution than the corresponding affinity distributions based on the isotherm analysis alone. We believe this will provide a useful tool for the study of the interactions by affinity biosensors.

APPENDIX A: RATE CONSTANT MAP THEORY

First, let us discuss the kinetics for biosensors. Consider the "1-to-1" kinetic model for the binding process. For each interaction we have

$$[A] + [L] \xrightleftharpoons[k_d]{k_a} [AL],$$
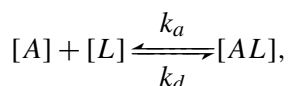
where k_a (k_d) denotes the association (dissociation) rate constant, and $[A]$, $[L]$, and $[AL]$ represent the concentrations of the analyte, ligand, and complex, respectively. In this work, the analyte A is injected and flushed over the surface in such a way that the concentration $[A](t)$ can be assumed to be constant during the study. The amount of free ligand will decrease with time according to $[L](t) = [L](0) - [AL](t)$. Suppose that the sensor response R is proportional to the complex concentration $[AL](t)$, that is, $R(t) = \lambda \cdot [AL](t)$, where λ is a constant. Denote by $R_{\max} = \lambda \cdot [L](0)$. The rate of complex formation will be

(19)
$$\frac{dR(t)}{dt} = k_a \cdot [A](t) \cdot (R_{\max} - R(t)) - k_d \cdot R(t)$$

assuming that the mass transfer kinetics are infinitely fast. Set $C = [A](t)$ (C is a constant, as mentioned above) and $R(t_0) = 0$, the solution to (19) is

(20)
$$R(t) = R_{\max} \cdot \frac{k_a C}{k_d + k_a C} \cdot \left(1 - e^{(k_d + k_a C)(t - t_0)}\right).$$

Now let us develop a *Rate Constant Map Theory* in the biosensor system. Assume that the binding can be described by an "m-to-n" kinetic model, that is, we have m analyte and n binding sites on the biosensor surface and first order kinetics. Denote by $(k_{a,i}, k_{d,j})$ the pair of association and dissociation constants for the interaction between the ith analyte and jth binding site. Let $R_{i,j}(t)$ be the response at time t of a complex with association constant $k_{a,i}$ and dissociation constant $k_{d,j}$.

Then, according to (20) we have

$$(21) \quad R_{i,j}(t) = \begin{cases} 0 & t \leq t_0 + \Delta t, \\[2mm] R_{i,j}^{\max} \dfrac{k_{a,i}C}{k_{d,j} + k_{a,i}C}(1 - e^{-(k_{d,j}+k_{a,i}C)(t-t_0)}), \\[2mm] \quad t_0 + \Delta t < t \leq t_0 + t_{inj} + \Delta t, \\[2mm] R_{i,j}^{\max} \dfrac{k_{a,i}C}{k_{d,j} + k_{a,i}C}(1 - e^{-(k_{d,j}+k_{a,i}C)t_{inj}})e^{-k_{d,j}(t-t_0-t_{inj})}, \\[2mm] \quad t > t_0 + t_{inj} + \Delta t, \end{cases}$$

where constant C is the concentration of the analyte, t_0 is the time when the injection of the analyte begins, and t_{inj} is the injection time. The adjustment parameter Δt is a time delay that accounts for the fact that it usually takes some time for the detector to respond to the injection. Constant $R_{i,j}^{\max}$ is the total surface binding capacity, corresponding to association and dissociation constants $k_{a,i}$ and $k_{d,j}$, that is, the detector response when every binding site on the biosensor surface has formed a complex with the analyte.

We now use the functions $R_{i,j}$ to make an approximation of the measured sensorgrams R_{obs}, by assuming that the total measured response, R_{obs}, of a system can be written as a linear combination of some individual responses, namely $R_{\text{obs}} = \sum_{i,j=1}^{m,n} R_{i,j}$. If we let $m, n \to +\infty$ in the above equation, we get the integral equation (1). Finally, we remark that the function $f(k_a, k_d)$, which is the generalization of the total surface binding capacity $\{R_{i,j}^{\max}\}$, is known as the (continuous) rate constant map. See Svitel et al. (2003) for details.

APPENDIX B: PROOF OF THEOREM 2

Since $\sigma_{\mathbf{c}}^2$ obeys the inverse Gamma distribution during the iteration by the definition of $\sigma_{\mathbf{c},k}^{-2}$ in (13), we have the uniform boundedness for the sequence $\{\sigma_{\mathbf{c},k}^{-2}\}$, that is (note that $(\mathbf{Lc})^T(\mathbf{Lc}) \equiv \|\mathbf{Lc}\|^2 \geq 0$),

$$0 \leq \sigma_{\mathbf{c},k}^{-2} = \frac{\alpha_{\mathbf{c}} + \frac{n}{2}}{\beta_{\mathbf{c}} + \frac{1}{2}\mathbb{E}_{q^k(\mathbf{c})}[(\mathbf{Lc})^T(\mathbf{Lc})]} \leq \frac{\alpha_{\mathbf{c}} + \frac{n}{2}}{\beta_{\mathbf{c}}}.$$

By the same argument, parameters $\{\sigma_{k,j}^2, j = 1, \ldots, N_C\}$ (note that $\sigma_{k,j}^{-2} := [\Sigma_k^{-1}]_{j,j}$) are also uniformly boundedness during the iterations. Therefore, a subsequence exists, denoted by $\{\sigma_{\mathbf{c},k_i}^{-2}, \sigma_{k_i,j}^2\}$ and fixed parameters $\{\sigma_{\mathbf{c},*}^{-2}, \sigma_{*,j}^2\}$ such that

$$\lim_{i\to\infty} \sigma_{\mathbf{c},k_i}^{-2} = \sigma_{\mathbf{c},*}^{-2}, \qquad \lim_{i\to\infty} \sigma_{k_i,j}^2 = \sigma_{*,j}^2, \quad j = 1, \ldots, N_C.$$

Since $\mathbf{c}_k$ and $\Sigma_k^{\mathbf{c}}$ are solely determined by parameters $\{\sigma_{\mathbf{c},k_i}^{-2}, \sigma_{k_i,j}^2\}$, we can deduce that $\lim_{i\to\infty} q^{k_i}(\mathbf{c}) = q^*(\mathbf{c})$. Using the formula (13), we can conclude $\lim_{i\to\infty} q^{k_i}(\sigma_{\mathbf{c}}^2) = q^*(\sigma_{\mathbf{c}}^2)$ and $\lim_{i\to\infty} q^{k_i}(\sigma_j^2) = q^*(\sigma_j^2), j = 1, \ldots, N_C.$

Now let us show that $q^*(\bullet) := q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^*(\sigma_{\mathbf{c}}^2)$ is the stationary point of the KL distance functional $\mathcal{D}$. By the definition of subsequence $\{\sigma_{\mathbf{c},k_i+1}^{-2}, \sigma_{k_i+1,j}^2\}$ in (13), subsequence $\{\sigma_{\mathbf{c},k_i+1}^{-2}, \sigma_{k_i+1,j}^2, j = 1, \ldots, N_C\}$ as well as subsequence $\{q^{k_i+1}(\mathbf{c}), q^{k_i+1}(\sigma_{\mathbf{c}}^2), q^{k_i+1}(\sigma_j^2), j = 1, \ldots, N_C\}$ also converge. Denote by

$$\lim_{i\to\infty} q^{k_i+1}(\mathbf{c}) = q^{**}(\mathbf{c}), \qquad \lim_{i\to\infty} q^{k_i+1}(\sigma_{\mathbf{c}}^2) = q^{**}(\sigma_{\mathbf{c}}^2),$$

$$\lim_{i\to\infty} q^{k_i+1}(\sigma_j^2) = q^{**}(\sigma_j^2).$$

Obviously, the following inequalities hold true

$$\mathcal{D}(q^{**}(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R}))$$

$$\leq \cdots$$

$$(22) \qquad \leq \mathcal{D}(q^*(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R})) \leq \cdots$$

$$\leq \mathcal{D}(q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R}))$$

$$\leq \mathcal{D}(q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^*(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R})).$$

Let us show that $q^*(\bullet) = q^{**}(\bullet) := q^{**}(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)$. To this end, define π as the algorithmic map of scheme (12), that is, the solution operator between the two iterations that maps $q^k(\bullet)$ into $q^{k+1}(\bullet)$. The continuity of the functional $\mathcal{D}$ implies the closedness of the mapping π. Hence, we have $q^{**}(\bullet) = \pi q^*(\bullet)$. This equation, together with inequalities (22) and the monotone convergence of $\mathcal{D}$, implies

$$\mathcal{D}(q^{**}(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R}))$$

$$= \mathcal{D}(q^*(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R})) = \cdots$$

$$(23) \qquad = \mathcal{D}(q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R}))$$

$$= \mathcal{D}(q^*(\mathbf{c})q^*(\sigma_1^2) \cdots q^*(\sigma_{N_C}^2)q^*(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R})).$$

On the other hand, it holds true that for any $q(\mathbf{c})$

$$\mathcal{D}(q^{**}(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R}))$$

$$(24) \qquad \leq \mathcal{D}(q(\mathbf{c})q^{**}(\sigma_1^2) \cdots q^{**}(\sigma_{N_C}^2)q^{**}(\sigma_{\mathbf{c}}^2)|p(\bullet, \mathbf{R})).$$

By combining the above inequality, the first inequality in (23), the first identity in (24), and the strict multi-convexity of the functional $\mathcal{D}$, we can deduce that $q^*(\mathbf{c}) = q^{**}(\mathbf{c})$. Furthermore, by repeating the preceding argument, we conclude that $q^*(\sigma_j^2) = q^{**}(\sigma_j^2)$ for $j = 1, \ldots, N_C$, and $q^*(\sigma_{\mathbf{c}}^2) = q^{**}(\sigma_{\mathbf{c}}^2)$, which completes the proof of identity $q^{**}(\bullet) = q^*(\bullet)$, and thus $q^*(\bullet)$ is the stationary point of $\mathcal{D}$. Finally, the relation (15) can easily be derived by applying $k_i \to \infty$ to the equations in (13).

Acknowledgements. We express our gratitude to the anonymous reviewers whose valuable comments and suggestions led to an improvement of the manuscript. The authors are grateful to Professor Mårten Gulliksson for the useful discussions. Cheng Li read and gave helpful comments on the paper. The authors would also like to thank Camilla Käck, Marie Andersson and Teodor Aastrup from our KK HÖG partner Attana AB for the PTH experiments. Professor Zhigang Yao is the corresponding author.

SUPPLEMENTARY MATERIAL

Supplement to "Estimating the rate constant from biosensor data via an adaptive variational Bayesian approach" (DOI: 10.1214/19-AOAS1263SUPP; .pdf). 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.

REFERENCES

ALTSCHUH, D., BJÖKELUND, H., STRANDGÅRD, J., CHOULIER, L., MALMQVIST, M. and ANDERSSON, K. (2012). Deciphering complex protein interaction kinetics using interaction map. *Biochem. Biophys. Res. Commun.* **428** 74–79.

ATHERTON, R., SCHAINKER, R. and DUCOT, E. (1975). On the statistical sensitivity analysis of models for chemical kinetics. *AlChE J.* **21** 441–448.

BABACAN, S. D., MOLINA, R. and KATSAGGELOS, A. K. (2008). Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Process.* **17** 326–339. MR2492977

BERNARDO, J.-M. and SMITH, A. F. M. (1994). *Bayesian Theory. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR1274699

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning. Information Science and Statistics*. Springer, New York. MR2247587

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776

BOX, G. E. P. and DRAPER, N. R. (1965). The Bayesian estimation of common parameters from several responses. *Biometrika* **52** 355–365. MR0210220

BROOKS, S., SMITH, J., VEHTARI, A., PLUMMER, M., STONE, M. et al. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde. *J. Roy. Statist. Soc. Ser. B* **64** 616–639.

CHEN, L., HOLST, M. and XU, J. (2009). Convergence and optimality of adaptive mixed finite element methods. *Math. Comp.* **78** 35–53. MR2448696

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

ELMLUND, L., KÄCK, C., AASTRUP, T. and NICHOLLS, I. (2015). Study of the interaction of trastuzumab and SKOV3 epithelial cancer cells using a quartz crystal microbalance sensor. *Sensors* **15** 5884–5894.

ENGL, H., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems*. Kluwer, Dordrecht.

EVANS, S. N. and STARK, P. B. (2002). Inverse problems as statistics. *Inverse Probl.* **18** R55–R97. MR1929274

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

GORSHKOVA, I., SVITEL, J., RAZJOUYAN, F. and SCHUCK, P. (2008). Bayesian analysis of heterogeneity in the distribution of binding properties of immobilized surface sites. *Langmuir* **24** 11577–11586.

GRAY, R. J. and PIERCE, D. A. (1985). Goodness-of-fit tests for censored survival data. *Ann. Statist.* **13** 552–563. MR0790556

GUPTA, A. and RAWLINGS, J. (2014). Comparison of parameter estimation methods in stochastic chemical kinetic models: Examples in systems biology. *AlChE J.* **60** 1253–1268.

JIN, B. and ZOU, J. (2010). Hierarchical Bayesian inference for ill-posed problems via variational method. *J. Comput. Phys.* **229** 7317–7343. MR2677781

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. and SAUL, L. (1999). Introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

KAIPIO, J. and SOMERSALO, E. (2007). Statistical inverse problems: Discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.* **198** 493–504. MR2260683

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398

KOMOROWSKI, M., COSTA, M., RAND, D. and STUMPF, M. (2011). Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc. Natl. Acad. Sci. USA* **108** 8645–8650.

KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York. MR0103557

LOH, W.-Y. and ZHENG, W. (2013). Regression trees for longitudinal and multiresponse data. *Ann. Appl. Stat.* **7** 495–522. MR3086428

PARISI, G. (1988). *Statistical Field Theory*. Addison-Wesley, New York.

PETERSON, C. and ANDERSON, J. (1987). A mean field theory learning algorithm for neural networks. *Complex Systems* **1** 995–1019.

PISCHEL, D., SUNDMACHER, K. and FLASSIG, R. (2017). Efficient simulation of intrinsic, extrinsic and external noise in biochemical systems. *Bioinformatics* **33** 1253–1268.

RIVARA, M.-C. (1984). Mesh refinement processes based on the generalized bisection of simplices. *SIAM J. Numer. Anal.* **21** 604–613. MR0744176

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR2080278

SANVICENS, N., MANNELLI, I., SALVADOR, J., VALERA, E. and MARCO, M. (2011). Biosensors for pharmaceuticals based on novel technology. *TrAC, Trends Anal. Chem.* **30** 541–553.

STANLEY, B. and GUIOCHON, G. (1995). Calculation of adsorption energy distributions of silica samples using nonlinear chromatography. *Langmuir* **11** 1735–1743.

STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785

SVITEL, J., BALBO, A., MARIUZZA, R., GONZALES, N. and SCHUCK, P. (2003). Combined affinity and rate constant distributions of ligand populations from experimental surface binding kinetics and equilibria. *Biophys. J.* **84** 4062–4077.

TELESCA, D., MÜLLER, P., KORNBLAU, S. M., SUCHARD, M. A. and JI, Y. (2012). Modeling protein expression and protein signaling pathways. *J. Amer. Statist. Assoc.* **107** 1372–1384. MR3036401

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TIKHONOV, A. N. and ARSENIN, V. Y. (1977). *Solutions of Ill-Posed Problems*. Winston & Sons, Washington, DC. MR0455365

WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

WAND, M. P., ORMEROD, J. T., PADOAN, S. A. and FRÜHRWIRTH, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Anal.* **6** 847–900. MR2869967

ZHANG, Y., FORSSÉN, P., FORNSTEDT, T., GULLIKSSON, M. and DAI, X. (2018). An adaptive regularization algorithm for recovering the rate constant distribution from biosensor data. *Inverse Probl. Sci. Eng.* **26** 1464–1489. MR3827396

ZHANG, Y., YAO, Z., FORSSÉN, P. and FORNSTEDT, T. (2019). Supplement to "Estimating the rate constant from biosensor data via an adaptive variational Bayesian approach." DOI:10.1214/19-AOAS1263SUPP.

ZIEGEL, E. and GORMAN, J. (1980). Kinetic modelling with multiresponse data. *Technometrics* **22** 139–151.

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the lasso. *Ann. Statist.* **35** 2173–2192. MR2363967

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

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