Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Bioinformatics and Computational Biology (CIBCB 2007)

Feature Sensitivity on Biochemical Signaling Pathways

George Papadopoulos and Martin Brown
The Control Systems Centre, School of Electrical and Electronic Engineering,
University of Manchester

Abstract – This paper investigates the solution of the feature selection problem in biochemical signal transduction pathways by examining the sensitivity of the features with respect to the model complexity using Basis Pursuit Regularization (BPR). Feature selection is effectively transformed into a continuous regularization problem with a characteristic 1-norm imposed on the parameter vector to penalize the models complexity. This technique makes possible the design of sparse models for the pathway data and because of the nature of the 1-norm it is possible to analyze the entire solution path (parameter locus) as the regularizer changes from zero to infinity.

I. INTRODUCTION

A major challenge in studying networks of cellular processes is the complex interactions among the genes, proteins or other compounds involved. The aim is to understand the nature of those interactions and consequently obtain some insight into the behavioral characteristics of the cell itself. Since the model of the process governing the cell is most of the times available, parameter estimation (and subsequently feature selection) is important in determining how exactly the various compounds interact. In the Nuclear Factor kappa Beta (NF-kB) cellular signaling pathway for example, the NF-kB is responsible for regulating numerous genes that are important for further inter and intra – cellular signaling, cell growth or apoptosis, making its precise understanding crucial [2].

Although the model is known there is a need to detect exactly how the various proteins or enzymes interact and this is largely controlled by the parameters in the model. Feature selection will inevitably lead into a reduced representation of the pathway and based on the parameters selected further identification of the importance of each compound will follow. The importance of NF-kB for example ([1][2]), as a drug target can now be exploited since it has been identified as a key in chronic inflammatory and auto immune diseases [2].

Previous work on feature selection in Systems Biology problems involved selecting the appropriate feature set based on parametric sensitivity analysis estimates [1] [2] [4] taking into account the oscillatory behavior of the systems output as it progressed through time. Although the results were very promising obtaining a final set of 9 features out of the total 64 (in the best case in [2], experiments were made in the NF–kB signaling pathway in both [2] and [1]), there is no consideration for possible colinearities between the variables that might allow for a case of a variable that is initially contributing positively in the estimation process changing attitude after the addition of another (Simpson's Paradox [13]). The novelty is largely on the study of the oscillations (of periods and magnitude) of the concentration of NF-kB and their use in deriving the sensitivity derivatives.

In [1] there is an attempt to cover the co-dependency of the parameters but based largely on [2] the reduced 9 parameter model is considered and the pair wise interactions were studied. Although providing promising results the questions that still remain is first, how the remaining parameters could have affected the original selection of the 9 used and second, how the remaining parameters could have affected the further dual modulation carried out, and of course the argument continuous recursively. A viable solution would be not to only employ pair wise modulation and examine the co-dependency from the start.

Most of the methods suffer from exactly that, none is emphasizing enough the fact that a variable might change attitude when another is added to the set. Discriminating features based on statistical measures such as t-test, the Fisher criterion [6][9][10][11], chi-test [10], Entropy feature ranking and PCA [10] [11][14] is hampered by the inherent assumption of normality (for optimal results) of the measures. Many of the methods act just as a preliminary stage towards classification and the bulk of the weight when achieving optimum performance is imposed on the classifier leaving the feature selection stage somewhat unattended. In addition, there are no objective criteria in selecting the optimal set or where to stop when ranking the features, there is no optimal rule for where to set the threshold other than trial and error.

Basis Pursuit deals with a continuous optimization problem instead of a discrete search one, yielding a globally optimum result with the colinearity problem addressed by making use of the piecewise linear trajectory of the parameter progression as the regularizer varies from zero to infinity. This is a way of studying the sensitivity of the feature selection process as the regularizer changes its value. This property allows researchers to study the whole parameter locus examining how each feature will behave as the whole family of sparse classifiers is calculated.

In the rest of the paper section II describes the model under consideration and gives the basics of its internal mechanism. Section III analyses the BPR method as a feature sensitivity tool outlining the importance of feature selection, what is meant by parameter activity and inactivity and how to obtain and use the parameter locus as a feature sensitivity tool with respect to the models complexity. Section IV is focused on the implementation of BPR in Ordinary Differential Equation (ODE) models as well as describing two commonly used techniques to work with the model in a more convenient way, namely the derivative and the integral approach, as well as

giving a detailed explanation as to why one preferred instead of the other. Section V includes the results obtained after applying BPR to the Raf Kinase Inhibitor Protein (RKIP) regulated on the Extracellular signal Regulated Kinase (ERK) pathway along with an analysis of the results of the feature selection process.

II. THE RKIP REGULATED ERK SIGNALING PATHWAY

Pathway models in biology are representations of the biochemical reactions governing the cell and are formulated as a set of differential equations describing these interactions. The models mostly used are linear in their parameters which represent the kinetic constants or reaction rates with which each reactant is participating in the biochemical process. The model consists of a set of ODEs assuming that spacial localization is not important and diffusion effects are not considered explicitly [15] (if this was the case then Partial Differential Equation (PDE) models would be appropriate [15]). In addition, it is also assumed that there is substantial number of molecules of each substrate that participate in the reaction since differential equation models represent by definition averages [15]. Alternatively, if this was not the case (*small copy number* [15]) then stochastic modeling would be more appropriate [15].

In general biochemical models can be non linear in the parameters (such as models describing the contact inhibition of Microcarrier Cultures [7] etc) suggesting non linear methods of optimization to be used. The work presented here will focus on cases where the system is linear in the parameters.

For the simulation and testing of the validity of BPR the Raf Kinase Inhibitor Protein (RKIP) on the Extracellular signal Regulated Kinase (ERK) signaling pathway model is considered. Although the detailed specifics are mentioned in [5] a brief introduction will be included here as well.

A. Model Description

The RKIP regulated ERK model shown in Figure 1 and used in subsequent analysis is somewhat restricted in that it only represents the ERK pathway regulated by RKIP. In [5] six events were identified while here only the ones represented in the schematic (relevant to our analysis) are presented: 1) As the pathway evolves through time Raf-1 (inhibited by RKIP) binds with RKIP with rate constants θ1 and θ2 to form the compound Raf-1/RKIP 2) free Raf-1 phosphorylates MARK/ERK Kinase (MEK) and inactive MEK is converted into active MEK-PP. This binds to ERK and phosphorylates it into ERK-PP. ERK-PP interacts with Raf-1/RKIP complex and forms Raf-1/RKIP/ERK-PP, 3) RKIP-P is released from the complex Raf-1/RKIP/ERK-PP with rate constant θ5 and binds with RP with rate constants θ θ9/ θ10 respectively to form the complex RKIP/RP. The RKIP-phosphotase (RP) is artificially introduced here to complete the model and emphasize the dephosphorylation of RKIP-P into the original active RKIP.

Previous work on this example [5] involved simple linearization of the difference equations and then solving for the parameters (since they enter in a linear fashion). Other approximations made to evaluate the derivatives on the LHS were based on polynomial interpolation of the simulated curves [5]. While this approach of approximation is valid theoretically it is very sensitive to noise which is increased in magnitude as the powers of the polynomial increase. The same effect on noise is true for the parameter estimation scheme where effectively if there is noise in the model it is considered to be normally distributed. The experimenter has to take a large number of estimates in time to combat this problem and in practice it is rarely used. In addition there is no feature selection stage/mechanism to qualitatively infer on the parameters. The sensitivity analysis is restricted in understanding the effects various compounds have on RKIP and ERK-PP concentration.

$$
\begin{aligned}
\frac{dx_1(t)}{dt} &= -\theta_1 x_1(t)x_2(t) + \theta_2 x_3(t) + \theta_5 x_4(t)\\
\frac{dx_2(t)}{dt} &= -\theta_1 x_1(t)x_2(t) + \theta_2 x_3(t) + \theta_{11} x_{11}(t)\\
\frac{dx_3(t)}{dt} &= \theta_1 x_1(t)x_2(t) - \theta_2 x_3(t) + \theta_3 x_3(t)x_9(t) + \theta_4 x_4(t)\\
\frac{dx_4(t)}{dt} &= \theta_3 x_3(t)x_9(t) - \theta_4 x_4(t) - \theta_5 x_4(t)\\
\frac{dx_5(t)}{dt} &= \theta_5 x_4(t) - \theta_6 x_5(t)x_7(t) + \theta_7 x_8(t)\\
\frac{dx_6(t)}{dt} &= \theta_5 x_4(t) - \theta_9 x_6(t)x_{10}(t) + \theta_{10} x_{11}(t)\\
\frac{dx_7(t)}{dt} &= -\theta_6 x_5(t)x_7(t) + \theta_7 x_8(t) + \theta_8 x_8(t)\\
\frac{dx_8(t)}{dt} &= \theta_6 x_5(t)x_7(t) - \theta_7 x_8(t) - \theta_8 x_8(t)\\
\frac{dx_9(t)}{dt} &= -\theta_3 x_3(t)x_9(t) + \theta_4 x_4(t) + \theta_8 x_8(t)\\
\frac{dx_{10}(t)}{dt} &= -\theta_9 x_6(t)x_{10}(t) + \theta_{10} x_{11}(t) + \theta_{11} x_{11}(t)\\
\frac{dx_{11}(t)}{dt} &= \theta_9 x_6(t)x_{10}(t) - \theta_{10} x_{11}(t) - \theta_{11} x_{11}(t)
\end{aligned}
\qquad (1)
$$

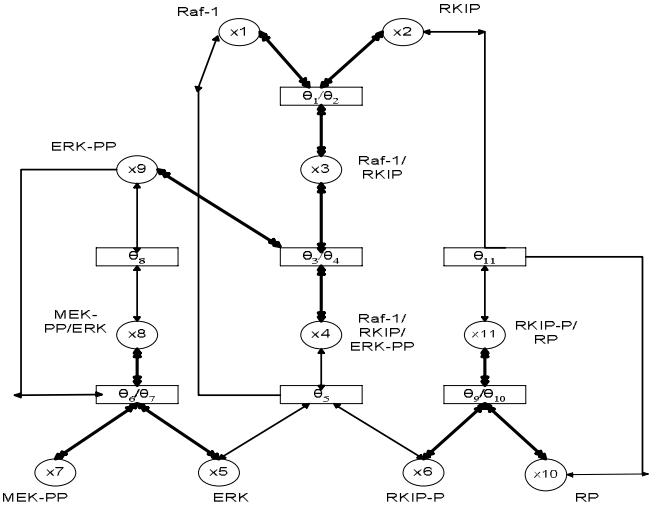

Figure 1 schematic representation of the RKIP regulated ERK signalling pathway. The rectangles represent the rate constants with which the various compounds (concentration in circles) interact. Solid arrows show bidirectional flow and thin ones single directions.

The argument posed in this paper is that BPR could provide a coherent and robust analysis of the behavior nor only of two

but all of the states in the model (as a result of the global study on feature selection) in relation to the parameters. Feature selection and as a consequence the study of its sensitivity with respect to the model complexity is therefore important in understanding exactly how RKIP affects the ERK pathway progression not necessarily based on sensitivity observations of the changes of the compounds concentrations but based on the parameters as well. In (1) and Figure 1 each parameter is with θ_i and each state (or circles in Figure 1) with x_i with the model having 11 states and 11 parameters.

III. FEATURE SENSITIVITY WITH BASIS PURSUIT REGULARIZATION.

One of the most important problems when selecting variables to include in a model is the possible interaction between them and specifically in biochemical systems, how they affect the further interpretation of parameter significance. This section deals with the basics of the BPR method, how it is derived and how the parameter locus is used to interpret feature importance.

A. The Importance of Feature Selection and Identifiability

Feature selection in general plays a very important role in any regression scenario and it is basically a process where the designer selects which features/variables will benefit the performance of the system more improving the identifiability of the system. The goal is to clean the data from irrelevant features (due to redundancy or noise for example) in order to make subsequent analysis easier either for regression or for understanding the general structure of the data. Feature selection is largely an empirical process in that sometimes the designer has to rely on prior knowledge of the system, in general this is a discrete non convex NP complete problem and so it is not feasible to examine every possible set of features and suboptimal techniques need to be applied like Sequential Forward Selection and variants where the optimal feature set is constructed in a sequential fashion by adding the most informative features one at a time. BPR differs in that instead of a combinatorial search for the optimal feature set the optimization of an objective function is sought. One important aspect is that it provides a globally optimum solution and the feature selection problem is now a continuous optimization problem which although piecewise quadratic it can be converted into a full quadratic programming problem which makes the solution tractable [3][8].

B. Basis Pursuit Regularization

In BPR feature selection is approached as a continuous optimization problem where the same formulation could be used to provide the solution for either a regression or a more general modeling problem. The important aspect of the algorithm is the use of a 1-norm on the parameter vector realizing in effect a penalty for the model complexity. This has the additional effect of introducing derivative discontinuities into the regularization function (when a parameter $\theta_i = 0$) that forces some of the parameters to be zero (sparseness into the parameter space) [3][8][12].

The parameter locus for the whole family of sparse classifiers can be computed since the evolution of the optimal parameter vector is a linear function of the regularizer. Feature selection can now be addressed by studying each parameters progression as the regularizer changes from zero (Maximum Likelihood solution) to infinity (all parameter values are zero).

Basis pursuit regression was proposed as a way of introducing sparseness into the feature space due to derivative discontinuities that appear when $\theta=0$ (since the derivative of the objective function cannot be defined). The objective function to be minimized is given by:

$$f(\boldsymbol{\theta}, \lambda, \mathbf{D}) = \tfrac{1}{2}\left\|\mathbf{t}-\mathbf{y}\right\|_2^2 + \lambda\left\|\boldsymbol{\theta}\right\|_1 \qquad (2)$$

where $\mathbf{y}$ is the m-dimensional model output vector, $\boldsymbol{\theta}$ is the n-dimensional parameter vector, λ is the non-negative regularization coefficient and $\mathbf{D}$ is the data set $\{\boldsymbol{\Phi},\mathbf{t}\}$. This objective is used directly with linear regression models:

$$y = \boldsymbol{\varphi}(\mathbf{x})^T\boldsymbol{\theta}$$

where the bias term associated with regression models is not present since it is assumed that the target vector and input features are all zero mean. The regression model assumes independent and identically distributed, zero mean, normally distributed measurement noise, where a Laplace prior distribution is placed on the parameter vector.

An important aspect of the basis pursuit regression approach is that it implicitly performs feature selection in a globally optimal fashion. It can be easily seen that the solution to (2) is unique and globally optimal, given a sufficiently varied training set.

To see the properties of BPR and to help understand why it implicitly performs feature selection, consider the differentiation of (2) with respect to the parameter vector. This gives:

$$\begin{aligned} &\boldsymbol{\Phi}^T\boldsymbol{\xi} = \lambda\,\mathrm{sgn}(\boldsymbol{\theta}) \\ &\boldsymbol{\theta} = \mathbf{H}^{-1}\left(\boldsymbol{\Phi}^T\mathbf{t} - \lambda\,\mathrm{sgn}(\boldsymbol{\theta})\right) \end{aligned} \qquad (3)$$

where $\boldsymbol{\theta}$ refers to the subset of active parameters, $\mathbf{H}=\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ and $\boldsymbol{\xi} = \mathbf{t}-\mathbf{y}$ and is termed the residual vector. While this isn't a direct solution, due to the presence of sgn($\boldsymbol{\theta}$) on the right hand side, it shows that the normal equations are similar to the least squares normal equations, apart from the extra "shrinkage" term. This partially shows the linear dependence of the parameter vector to the regularizer λ, the second part of (3) is obtained after solving for the optimal parameter vector $\boldsymbol{\theta}$. Equations (3) are not necessary for the application of the method; they are simply presented here to illustrate mathematically the piecewise linearity of the parameter vector trajectory. Although these equations show that the parameter vector moves linearly along various segments what remains is to further explain how these segments are defined. From (3) by removing the sign we have:

$$-\lambda\mathbf{1} \le \boldsymbol{\Phi}^T\boldsymbol{\xi} \le \lambda\mathbf{1} \qquad (4)$$

which defines a (-λ,λ) box within which all parameters should optimally reside. The change in the set of parameters entering and sometimes exiting the box (Simpson's Paradox) defines the knots in the piecewise linear curve.

C. Parameter Activity and Inactivity

The box defined by (4) depicts an implicit set of active/inactive parameters at each iteration of the algorithm. The inactive parameters are defined as simply the ones that are contained within:

$$-\lambda \mathbf{1} < \mathbf{\Phi}^T \boldsymbol{\xi} < \lambda \mathbf{1} \qquad (5)$$

while the active ones lie in the boundary or outside the box. The realization of activity/inactivity plays an important role in the efficient calculation of the parameter locus and is linked with the piecewise nature of the parameter trajectory. Furthermore, the active parameters will have non zero value or $|\boldsymbol{\theta}| > 0$ and the inactive ones $|\boldsymbol{\theta}| = 0$, it can also be shown that by considering small perturbations of the inactive parameters the inner product between their features and the residuals is bounded above by λ. $\mathbf{\Phi}^T \boldsymbol{\xi} = \lambda \operatorname{sgn}(\boldsymbol{\theta})$ gives a set of n_A equality constraints that the active parameters must satisfy and (5) a set of $2n_I$ inequality constraints for the inactive ones. The change from active to inactive and vice versa signifies a knot in the parameter locus and the feature selection capability lies in studying the locus and selecting the appropriate λ associated with the corresponding step in the algorithm progression.

D. Regularization Parameter Locus

One important feature for this class of regression models is that the parameter locus $\boldsymbol{\theta}(\lambda)$ is a piecewise linear curve, stretching from $\boldsymbol{\theta}(\infty) = 0$ to the maximum likelihood solution at $\boldsymbol{\theta}(0)$ (assuming that H is non-singular). The "knots" in the piecewise linear locus occur when a parameter changes state from inactive (in the box defined by(-λ,λ)) to active (outside or equal to λ), or vice versa, and the dimension of the optimal solution in (3) changes. When λ is locally perturbed and the active parameter set does not change, the normal equations in (3) demonstrate that the parameters are a linear function of λ, so the optimal parameter locus around that point is linear. It is possible to efficiently compute the complete optimal parameter locus by starting at $\boldsymbol{\theta} = 0$ and this is discussed further in [3][12].

Feature selection is based upon studying the behavior of the progression of the parameters as the regularizer λ changes. Highly correlated or noisy features will in generally drop out in early stages of the parameters locus progression. In addition, in order to make a decision for whether a parameter aids the performance of the model or not it is important to see how the errors progress as λ changes. It is then a straight forward task to infer on a parameters importance to the model. A more detailed description of the method will be presented in the section V where BPR is applied in the Michaelis-Manten signaling pathway.

E. Parameter Locus Example

From (3) it is evident that the parameter locus is piecewise linear as λ changes from zero to infinity. To illustrate this effect graphically a simple example was chosen depicting the Homogeneous Gas Phase Reaction of NO with O2 (the Bodenstein – Linder model) :

$$2NO + O_2 \leftrightarrow 2NO_2$$

The ODE model is given by:

$$\frac{dx}{dt} = \theta_1 (a - x)(b - x)^2 - \theta_2 x^2$$

where θ_i, i=1,2 are the parameters and a, b are constants given by 126.2 and 91.9 respectively.

The example was chosen so as to illustrate the Simpson's Paradox [13]effect where the addition of an extra parameter may cause a previously included one to drop out of the optimal feature set. Figure 2 shows the parameter locus generated where it is clear from the early stages that although θ_1 has initially a beneficial effect the addition of θ_2 causes it to temporarily drop out until the regularizer changes value accordingly.

Figure 3 indicates how θ_1 changes with respect to θ_2. Notice how this version of the locus changes sign from positive to negative showing the negative effect that one parameter has to the other as the regularizer changes value.

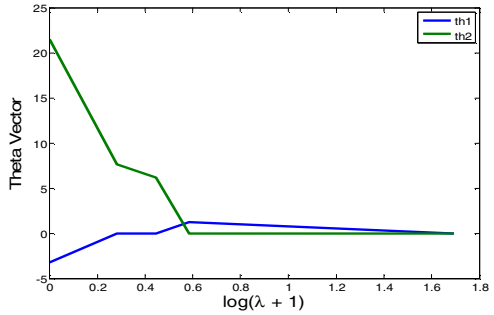

Figure 2 The piecewise linear trajectory illustrated for two parameters as the regularizer was varied from zero to infinity.

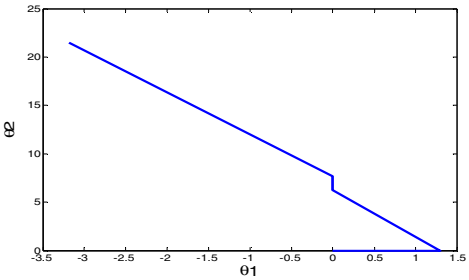

Figure 3 The parameter locus showing the changes of one parameter with respect to the other.

As the regularizer changes from zero to infinity the parameter locus can give a detailed picture of how the feature selection process is affected by the model complexity, the

appropriate model can now be selected based on information about the parameters individual behaviour and in relation to others.

IV. APPLYING BPR TO ODE MODELS.

In the previous section the basic theoretical aspects were covered leaving now the implementation part of the problem to be addressed. When all the states are measurable and when the parameters enter the model in a linear fashion one can often use shortcut methods to reduce the parameter estimation problem from an ODE system to an algebraic one [7].

Consider the general model formulation presented in (6) where the system now has *m* states and *n* parameters with $\boldsymbol{x}$ being the states and $\boldsymbol{\theta}$ the parameters. A more compact representation in matrix format is given in (7) where $\dot{\mathbf{x}}$ is the derivative of the states (structures as a vector), Φ is an $m \times n$ matrix (*m* states and *n* variables) formed by the combinations of the states and $\boldsymbol{\theta}$ the parameter vector and y is the output vector where all the states are measurable and it is assumed that there is corruption from a zero mean and σ^2 variance random sequence.

$$\begin{aligned}
\frac{dx_1}{dt} &= \theta_1\varphi_{11}(x) + \theta_2\varphi_{12}(x) + \ldots + \theta_n\varphi_{1n}(x) \\
\frac{dx_2}{dt} &= \theta_1\varphi_{21}(x) + \theta_2\varphi_{22}(x) + \ldots + \theta_n\varphi_{2n}(x) \\
&\vdots \\
\frac{dx_m}{dt} &= \theta_1\varphi_{m1}(x) + \theta_2\varphi_{m2}(x) + \ldots + \theta_n\varphi_{mn}(x)
\end{aligned} \tag{6}$$

$$\begin{aligned}
\dot{\mathbf{x}} &= \Phi(\mathbf{x}, t)\boldsymbol{\theta} \\
\mathbf{y} &= \mathbf{x} + N(0, \sigma^2)
\end{aligned} \tag{7}$$

Notice that this is still the original ODE dynamical system with the main problem being the derivatives on the LHS of both (6) and (7). The task is now to either approximate them or see if by integrating both parts (since there is access to the original state curves) could lead in converting the dynamic system to a static, linear in the parameters model representing the behavior of the system from time $t_0 = 0$ to time t_i.

A. The Derivative Approach

Consider the state-space model formulation as given in (6) which depicts the general structure for a model having m states and n parameters covering the case where the parameters enter in a linear fashion. The problem with this set of ODEs is the derivatives on the LHS of the equation, if these were to be calculated then the system would have the form:

$$\begin{aligned}
\underline{\eta} &= \left[\frac{d\hat{\mathbf{x}}(t)}{dt}\right]_{t=t_i} \\
\underline{\eta} &= \Phi(\hat{\mathbf{x}})\underline{\theta}
\end{aligned} \tag{8}$$

In the derivative approach the LHS derivative of (7) is approximated using an interpolated version of the state trajectories.

Based on (8) the BPR objective function can now be stated as:

$$J(\underline{\theta}) = \sum_{i=1}^{N}\left[\underline{\eta}_i - \Phi(\hat{\mathbf{x}}_i)\underline{\theta}\right]^T Q_i\left[\underline{\eta}_i - \Phi(\hat{\mathbf{x}}_i)\underline{\theta}\right] + \lambda\|\underline{\theta}\|_1 \tag{9}$$

where J is the objective function to be minimized, θ is the parameter vector, Q is a weighting matrix so as to cover the general case where the designer needs additional information about the statistical properties of the data to be incorporated into the solution, the subscript *i* denotes the time instants and in Q could signify the use of a different weighting at each point in time, in the present case no weighting was applied and so $Q = I$ for every time point.

Another important feature is the summation in time in (9); this is an important aspect associated with the derivative approach since it solves the singularity present when considering each single time instant (described in detail in subsection C).

To proceed in using (9) one needs to have an approximated curve fit of the state trajectories represented in the LHS of (6) constituting the targets in the objective function. Strictly speaking the data fit should be constructed in such a way so as to adequately represent the state trajectories and at the same time smooth out any random noise components [7], the derivative approach has been criticized as inaccurate since it actually amplifies the measurement noise present in the model, if the model used is indeed the true model describing the process governing the cell then an approximation as close as possible to the original curves (interpolation) is sufficient [7]. This is a trial and error procedure and in most cases a visual inspection of the fitted versus the original curves as well as an estimate of the normality of the residuals (this under the assumption that any measurement noise is at least zero mean) should be sufficient to give a qualitative estimate of the fitting process [7].

The next subsection describes another alternative for a shortcut estimation method for the parameters, namely the Integral approach where now (6) is integrated in both sides so as to obtain an expression based on the states constituting the solutions of ODE model.

B. The Integral Approach

Starting again from (6) and integrating both sides the ODE model is now transformed into:

$$\begin{aligned}
\int\frac{d\mathbf{x}}{dt}dt &= \int_{t_0}^{t_i}\Phi(\mathbf{x})dt\underline{\theta} \\
\mathbf{x}(t_i) &= \mathbf{x}_0 + \Psi(t_i)\underline{\theta}
\end{aligned} \tag{10}$$

where $\Psi(t_i) = \int_{t_0}^{t_i}\Phi(\mathbf{x})dt$ and integration takes place from time $t_0 = 0$ to time t_i, the objective function is now given by:

$$J(\underline{\theta}) = \sum_{i=1}^{N}\left[\hat{\mathbf{x}}_i - \mathbf{x}_0 - \Psi(t_i)\underline{\theta}\right]^T Q_i\left[\hat{\mathbf{x}}_i - \mathbf{x}_0 - \Psi(t_i)\underline{\theta}\right] + \lambda\|\underline{\theta}\|_1 \tag{11}$$

where again the *i* subscript denotes the different points in time and $Q = I$ in this scenario as well.

Although the integral approach will not in general be prone to noise amplification as it is the case with the derivative

approach, it will produce, in the current case (the Michaelis-Manten model), a singular data matrix resulting in a system with infinite number of optimal solutions (singularity is inherent because of the structure of the resulting algebraic model, more on the next subsection) making the derivative alternative much more reliable.

C. Singularity Analysis

To analyze the singularity present and see how the integral approach fails to provide adequate results a simpler pathway model is considered that suffers from the same problem as the ERK one. The Michaelis-Manten signaling pathway is a 4 state 3 parameter model represented in(12):

$$\begin{aligned}\frac{dE(t)}{dt} &= -\theta_1 E(t)S(t) + (\theta_2+\theta_3)ES(t)\\ \frac{dS(t)}{dt} &= -\theta_1 E(t)S(t) + \theta_2 ES(t)\\ \frac{dES(t)}{dt} &= \theta_1 E(t)S(t) - (\theta_2+\theta_3)ES(t)\\ \frac{dP(t)}{dt} &= \theta_3 ES(t)\end{aligned} \tag{12}$$

$$\dot{\mathbf{x}} = \Phi(\mathbf{x},t)\boldsymbol{\theta} \tag{13}$$

or

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\\ \dot{x}_4\end{bmatrix} = \begin{bmatrix}-x_1x_2 & x_3 & x_3\\ -x_1x_2 & x_3 & 0\\ x_1x_2 & -x_3 & -x_3\\ 0 & 0 & x_3\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\end{bmatrix}$$

where the state matrix **Φ** is rank deficient. To illustrate the co-dependency of the parameters it is sufficient to reach the low echelon form of **Φ** having at least one row of zeros. Indeed the low echelon form of **Φ** (its transpose is used here in order to show more clearly the dependence on the parameters) is given by:

$$\begin{bmatrix}-x_3x_1x_2 & -x_3x_1x_2 & x_3x_1x_2 & 0\\ 0 & 0 & 0 & 0\\ x_3 & 0 & -x_3 & x_3\end{bmatrix}$$

showing that there is a linear dependence between parameters θ1 and θ2 resulting in the state matrix having rank 2. The immediate meaning is that for a single instant in time the system is singular; for the solution to be unique one has to consider at the whole time span (at least two time intervals) so as not to allow the state matrix to loose rank, something that it is covered when summing in time using the derivative approach. With the integral approach however because of the integration action the singularity persists making the method unsuitable for use in this scope.

In addition, the state matrix includes nonlinear combinations of the states and in general there is no input driving the system (or assumed to be constant) and assuming sufficient smoothness of **Φ(x)** the model can be considered as a nonlinear (in the states) autonomous system. A natural way of reaching a solution would be to linearize about the equilibrium points resulting in a linear (in the states) state-space model. Close examination however reveals that there is an infinite set of equilibrium points obtained by $\Phi(\mathbf{x}) = 0$ because of the rank deficiency of **Φ(x)**, leading to infinite many $\nabla_x \Phi(\mathbf{x})$ Jacobean matrices after linearization. This reinforces the previous statement that for a single instant in time the problem is singular the solution to which is considering the whole time span of study.

The problem with the integral approach is now more apparent since there is integration in time entering the process which will result in a system being again singular.

V. RESULTS AND DISCUSSION

The RKIP regulated ERK signaling pathway was considered so as to test the feature selection sensitivity with respect to the modeling complexity capabilities of BPR. Figure 4 shows its simulation using the nominal parameter ([θ1, θ2, θ3] =[0.46, 0.016, 0.7]) values and initial conditions presented in Table 1 and for the estimation process normally distributed zero mean unit variance measurement noise was also introduced.

State	x1	x2	x 3	x 4	x 5	x 6
Init. Value	70.0010	47.6273	17.3763	22.3146	1.4151	13.3234
Nom. Param.	0.5300	0.0072	0.6250	0.0024	0.0315	0.8000

State	x 7	x 8	x 9	x 10	x 11
Init. Value	68.8972	87.2797	1.6715	46.1116	76.1367
Nom. Param.	0.0075	0.0710	0.9200	0.0012	0.8700

Table 1 Table summarizing the initial condition and nominal parameter values for RKIP regulated ERK signalling pathway.

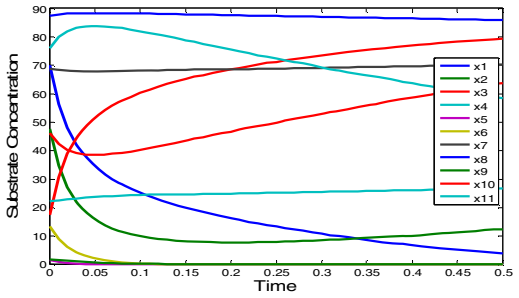

Figure 4 Simulation of the RKIP regulated ERK signalling pathway.

Proceeding in estimating the optimal parameter values using BPR it is evident (Figure 5) that BPR provides optimal results (exact parameter estimation) for a value of $\lambda = 0.01$. The crossed lines in the figure below represented the estimated curves.

When the estimation method is applied the system is been modified to include some added normally distributed zero mean noise so as to make the estimation process more realistic.

Figure 5 shows the estimation outcome of the system where the solid linea are the target curves and the estimated ones are represented by the corresponding crosses; Figure 6 shows the

parameter locus as λ changes. Analyzing the parameter locus further, although most of the parameters affect the progress of the system at a very early stage as λ progresses parameters 1, 9, 11, 6, and 4 appear to be the most significant (arranged in order of importance).

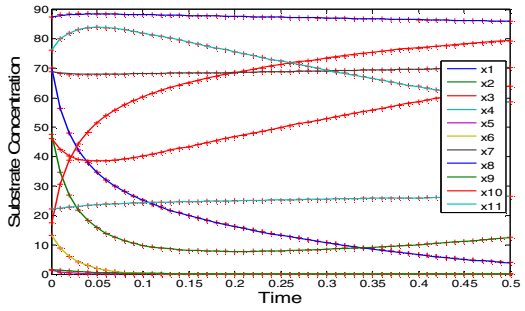

Figure 5 Simulation of the ERK pathway and the predicted curves estimated using the BPR method. Solid lines are the target curves and the crosses represent the estimated ones.

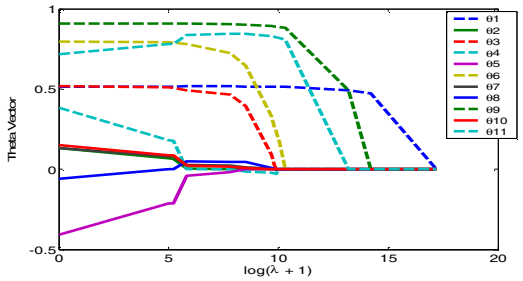

Figure 6 The parameter locus of the ERK model. Out of the 11 parameters only 5 exhibit substantial control.

A. ERK Pathway Model Analysis

Another key aspect is that features 2, 5, 7, 10, 8 never enter into the process and still the system reaches optimal performance. This is of great importance since now apart from a minimal model representation the importance of the remaining parameters (based on the changes of the error as λ varies, Figure 7) is also available. It is obvious that after the fifth parameter enters the process (observing Figure 7 from right to left) there is no significant change in the overall error of the system, signifying the fact that even if these parameters were omitted they would still cause no significant effect on the system. Figure 7 is to be read from left to right taking into consideration that the error is increasing as more parameters are left out of the model, so in this case even if 5 of the features are removed (the ones mentioned above) the error still remains very close to zero. As soon as more significant figures are removed the error is incresed (reaching a maximum at the right side of the curve).

The question that still remains is how this information can now be studied in conjunction with the model itself, what can be learnt from the feature selection process or how to assign meaning to the feature selection results.

Each parameter represents a kinetic/rate constant quantifying the rate with which a reactant is taking part into a biochemical reaction. When the parameter associated with a reactant is insignificant it is indicative that the reactant in question carries no substantial effect into the outcome of the model. In relation to the minimal model representation, if a parameter does not enter into the feature selection process at all it implies that the substance associated with the parameter has no effect to the outcome of the system at all. It is this interpretation of the modeling outcome that is providing the means to compartmentalize the signaling pathway model into different parts each providing different weight to the progress of the system.

Consider the schematic of the ERK model depicted in Figure 1, observing carefully, parameters (kinetic/rate constants) 1, 9, 11, 6, and 4 are located in separate parts of the network, each in one of the four previously identified as separate subsections of the whole pathway [5] signifying important events in the whole signaling network. It should be mentioned that previous studies [5] have analyzed and identified the mechanisms of the entire ERK pathway, there is no distinction of which one is important and which not or how the progress of the pathway will be affected by disregarding irrelevant features. The compartmentalized version of the pathway is presented in Figure 8 where the feature selection outcome is being translated into the pathway.

The results obtained from BPR are therefore not only in accordance with previous independent studies but also provide a further insight into the whole structure of the ERK pathway. Furthermore, this would mean that (Figure 8):

1. When the Raf-1/RKIP complex is formulated the Raf-1 substrate is combined with RKIP and the rate in which Raf-1 is introduced is of crucial importance.
2. This could also affect the free Raf-1 that will ultimately phosphorylate MEK and convert inactive MEK into active MEK-PP.
3. When MEK-PP is binding with ERK the rate of MEK-PP introduction is more important as MEK-PP/ERK is phosphorylated into active ERK-PP.
4. The RKIP-phosphotase (artificially introduced in the example) is less important in the formulation of RKIP-P/RP than RKIP-P. This also affects the feedback converting RKIP-P/RP back into RP as well as the generation of RKIP.
5. Parameter θ_{11} is very important since it is related to two separate events, affecting indirectly the rate with which RKIP is produced and the later production of RKIP which is combined with Raf-1. This is why θ_2 is not included into the optimal set; it is influenced indirectly by θ_{11}.
6. Parameter θ_4 controls the rate with which Raf-1/RKIP binds with ERK-PP to form the complex Raf-1/RKIP/ERK-PP and it is more important than the rate with which ERK-PP is introduced. The Raf-1/RKIP/ERK-PP complex is largely responsible for the release of active Raf-1 which later increases the release of MEK-PP and eventually ERK-PP [5], so in a sense parameter θ_3 is linked with θ_4 and the loop seems to be

initiated and largely controlled by the rate of introduction of Raf-1/RKIP.

7. In simple terms, the minimal model (disregarding features 2,3,5,7,8,10) should be translated into if the control of the pathway is needed then most of the events could be controlled by only considering features 1,9,11,6 and 4.

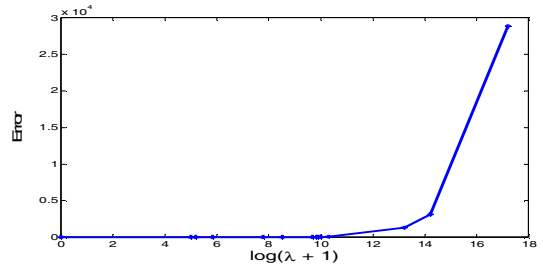

Figure 7 The error locus of the BPR method in the ERK model. As more features are added into the process the error of the system is reaching zero.

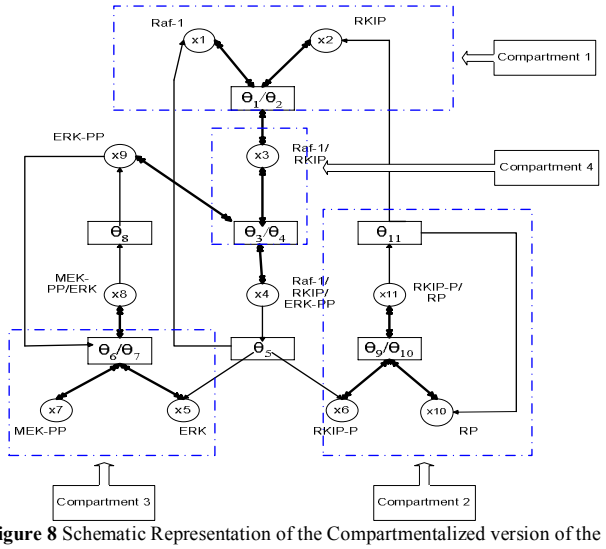

Figure 8 Schematic Representation of the Compartmentalized version of the ERK pathway. The most significant parameters lead to the version of the pathway shown.

VI. CONCLUSION

The Basis Pursuit Regularization (BPR) method has been applied to the RKIP regulated ERK signal transduction pathway where an analysis of the feature selection process with respect to the model complexity was performed. BPR addresses the feature selection process as a feature sensitivity with respect to the regularizer variable (model complexity) and the problem now is effectively a continuous optimization problem due to the 1-norm complexity penalty on the parameter vector, introducing sparseness into the parameter space. The nature of the norm implies a linear relationship of the parameter vector with the regularizer in the objective function, allowing the calculation of the whole family of sparse models (parameter locus). Feature selection can now be studied easily since the designer can see how the parameters change as the regularizer changes from zero (Maximum Likelihood Solution) to infinity (all the parameter are zero) and at the same time examine possible colinearities (Simpson's Paradox [13]) in the features. In the example considered the main scope of the analysis was to examine parameter interactions (model sensitivity) and how these affect the outcome of the model.

Future work will focus on how to implement the method in cases where there is no linear dependence on the parameters as well as assessing the sensitivity derivatives of the feature set derived from BPR. Combining BPR and normal methods in sensitivity analysis may provide a way to examine how to obtain an expression of the parameter locus for dynamic models while at the same time considering the possible colinearities between features.

VII. REFERENCES

[1]. Broomhead DS, Ihekwaba AEC, Grimley R, White MRH, Benson N, Kell D, " Synergistic control of oscillations in the NF-kB signaling pathway", Sep 2005, IEE Proc, Systems Biology, Vol 152, No 3, p153-160.

[2]. Broomhead DS, Ihekwaba, Grimley, Benson, Kell, "Sensitivity analysis of parameters controlling oscillatory signaling in the NF-kB pathway: the roles of IKK and IkBa". Systems Biology, vol 1 No 1, June 2004.

[3]. Brown M, N.P. Costen, "Exploratory Basis Pursuit Classification", Pattern Recognition Letters 2005, vol 26, no 12, p1907-1915.

[4]. Cho KW, Shin SY, Wolkenhauer O, Kolch W, "Experimental Design in Systems Biology, based on parameter sensitivity analysis using a Monte Carlo method: a case study for the TNFa – Mediated NF – kB signal transduction pathway", Dec 2003, Simulation, Vol 79, p726-739.

[5]. Cho KW, Shin SY, Kim HW, Wolkenhauer O, McFerran B, Kolch W, "Mathematical modeling of the influence of RKIP on the ERK signaling pathway",

[6]. Chuang. H,Liu. H, Brown S, Kao C, Hsu F, "Identifying Significant Genes from Microarray Data", 4th Bioinformatics and Bioengineering Symposium, 2004.

[7]. Englezos P, Kalogerakis N, "Applied parameter estimation for chemical engineers", 2001, Marcel Dekker.

[8]. Hastie T, Johnstone, Tibshirani, "Least Angle Regression", Technical Report, Department of Statistics, Stanford University, June 16 2003.

[9]. Li T, Zhang C, Ogihara M, "Acomparative study of feature selection and multiclass classification methods for tissue classification based on gene expression", 2004, Bioinfomatics, Vol 20, No 15, p2429-2437.

[10]. Liu H, J.Li, L.Wong, "A comparative study on feature selection and classification methods using Gene Expression profiles and proteomic patterns".Genome Informatics, vol 13, p51-60, 2002 .

[11]. Model F, Adorjan P, Olek A, Piepenbrock C, "Feature Selection for DNA methylation based cancer classification", Feb 2001, Bioinformatics, Vol 17, pS157-S164.

[12]. Rosset S, Zhu J, Hastie T, Tibshirani R, "1-norm Support Vector Machines", NIPS 2003 Workshop, http://clopinet.com/isabelle/Projects/NIPS2003/.

[13]. Westbrooke Ian, "Simpson's Paradox, an example in a New Zealand survey of jury composition", Research Report #1, Statistics New Zealand Te Tari Tatau, Aug 1997.

[14]. Yun Z, Keong K C, "identifying Simple Discriminatory Gene Vectors with an Information Theory Approach", Computational Systems Bioinformatics Conference, 2005.

[15]. Sontag E.D, "Molecular Systems Biology and Control", European Journal of Control, Vol.11, 2005, p1-40.