Sensitivity analysis of complex stochastic processes

Yannis Pantazis
join work with Markos Katsoulakis

University of Massachusetts, Amherst, USA
Partially supported by DOE contract: DE-SC0002339

February, 2013

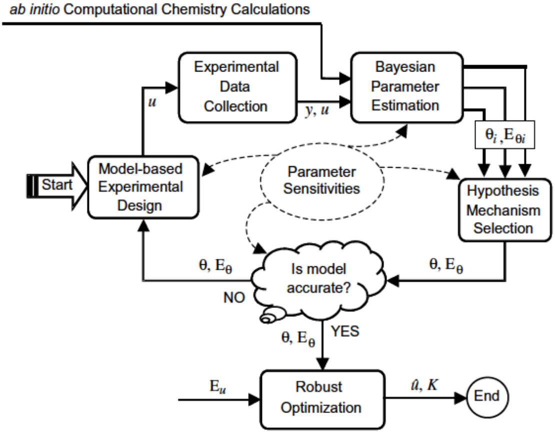
Outline

Sensitivity analysis - Definition

- Definition: Sensitivity analysis is the study of the impact in the output of a mathematical model or system caused by perturbations in the input.
- Q: What could be an input?
  A: Initial data (deterministic or random), intrinsic model parameters (such as temperature, pressure, masses, reaction rates, etc.)
- Q: What could be an output?
  A: Output data, observables of the process (such as mean values, variances, higher moments, correlations, etc.), histograms.
- An input parameter is sensitive when even small changes to this parameter may significantly alter the system output.

Sensitivity analysis - Applications

- *Perspectives on the design and control of multiscale systems*, Braatz et al., 2006



Sensitivity analysis - Applications

- Optimal experimental design: Construct experiments in such a way that the parameters can be estimated from the resulting experimental data with the highest statistical quality. Various optimality criteria (A-optimality, D-optimality, E-optimality, etc.) are based on the Fisher information matrix.
- Robustness: The persistence of a system to a desirable state under the (controllable and/or uncontrollable) perturbations of external conditions.
- Identifiability: The ability of the experimental data and/or the model to estimate the model parameters with high fidelity.
- Reliability: To ensure that the performance of a system meets some pre-specified target with a given probability.
- Uncertainty quantification: The science of quantitative characterization and reduction of uncertainties in applications. For instance, not only the values of the model parameters might contain errors but also the model itself could be an approximation.

Sensitivity analysis - Global vs Local

- Global sensitivity analysis: Study the effect of an input parameter to the output under a wide range of values or under a specified distribution. Insensitive input parameters could be treated without big effort (assign a nominal value) while sensitive input parameters should be treated carefully.
  - Variance-based methods for global SA are: Sobol' indices (decompose variance in a ANOVA-like manner), Fourier amplitude sensitivity test (FAST), Morris method.
  - Information-based methods for global SA are: relative entropy (or Kullback-Leibler divergence), mutual information (quantifies the input-output mutual dependence).
- Local sensitivity analysis: Fix input parameters to a value and study the effect of small perturbations around the fixed parameter to the output. Local SA is typically a prerequisite for global SA.

Outline

Deterministic SA

- System of ODEs:

$$\dot{y} = f(t, y; \theta) \quad , \quad y(0) = y_0 \in \mathbb{R}^N$$

  - Goal: Perform SA on the model parameters $\theta \in \mathbb{R}^K$.
- Define sensitivity indices:

$$s_k = \frac{\partial y}{\partial \theta_k}$$

- A new system of ODEs is derived and augmented to the previous:

$$\dot{s}_k = \frac{\partial f}{\partial y} s_k + \frac{\partial f}{\partial \theta_k}, \; k = 1, ..., K$$

Stochastic SA - Observable-based

- By stochastic process we merely mean either discrete-time Markov chains (DTMC) with possibly separable state space or continuous-time jump Markov chains (CTMC) with countable state space.
  - Well-mixed reaction networks, spatially-extended on lattices systems, discretizations of stochastic differential equations are models which belong to the above classes of stochastic processes.
- A typical example for stochastic SA:
  - $\{x_t\}$ is a MC, $\theta \in \mathbb{R}$ is a model parameter and f is a test function (or observable).
  - Compute

$$S(\theta, t) = \frac{\partial}{\partial \theta} \mathbb{E}_{P_t^\theta}[f(x)]$$

Stochastic SA - Observable-based

- Finite difference approach. Approximate parameter sensitivity with finite difference,

$$S(\theta, t) = \frac{\mathbb{E}_{P_t^{\theta+\epsilon}}[f(x)] - \mathbb{E}_{P_t^{\theta}}[f(x)]}{\epsilon} + O(\epsilon)$$

- Applying typical sampling approaches,

$$\hat{S}(\theta, t) = \frac{1}{n\epsilon} \sum_{i=1}^{n} \left(f(x_t^{\theta+\epsilon,i}) - f(x_t^{\theta,i}) \right)$$

- with variance,

$$\begin{aligned} \mathrm{var}\left(f(x_t^{\theta+\epsilon}) - f(x_t^{\theta}) \right) = & \ \mathrm{var}\left(f(x_t^{\theta+\epsilon}) \right) + \mathrm{var}\left(f(x_t^{\theta}) \right) \\ & - 2\mathrm{cov}\left(f(x_t^{\theta+\epsilon}), f(x_t^{\theta}) \right) \end{aligned}$$

- Importance quantities:
  - Bias due to finite difference.
  - Variance due to sampling of two different stochastic processes.

Stochastic SA - Observable-based

- Likelihood ratio method, P. Glynn (1987–90). Parameter sensitivity is rewritten as,

$$S(\theta, t) = \frac{\partial}{\partial\theta}\mathbb{E}_{P_t^\theta}[f(x)] = \int f(x)\partial_\theta P_t^\theta(x)dx$$
$$= \mathbb{E}_{P_t^\theta}[f(x)\partial_\theta \log P_t^\theta(x)]$$
$$= \mathbb{E}_{Q_{0,t}^\theta}[f(x)\partial_\theta \log Q_{0,t}^\theta(x)]$$

  - where $Q_{0,t}$ is the path distribution of the process in the interval $[0, t]$.
- Derivative-free approach.
- Variance increases with time.

Stochastic SA - Density-based

- A more holistic approach based on the probability densities of the process.
- Assuming that $x_t^\theta \sim P_t^\theta(x)dx$ and $x_t^{\theta+\epsilon} \sim P_t^{\theta+\epsilon}(x)dx$, why not comparing directly the distributions?
  - Majda and Gershgorin (2010) compared the distributions between a coarse-grained climate model and a statistical climate model based on relative entropy.
  - Komorowski et al. (2011) computed the Fisher information matrix of the path distribution of a linearized CTMC process.

Outline

Preliminaries

- Restrict, for the moment, to DTMC processes and to the steady state regime.
  - Steady state means that the probability density of the process an any time instant is independent of the particular time instant.
- For the parameter vector $\theta \in \mathbb{R}^K$, a DTMC, $\{\sigma_m\}_{m\in\mathbb{Z}_+}$, is defined with path distribution in time instants $0, 1, ..., M$ given by

$$Q^\theta_{0,M}(\sigma_0, \cdots, \sigma_M) = \mu^\theta(\sigma_0) p^\theta(\sigma_0, \sigma_1) \cdots p^\theta(\sigma_{M-1}, \sigma_M)$$

  where
  - $\mu^\theta(\sigma)$ is the stationary (or invariant) probability density function.
  - $p^\theta(\sigma, \sigma')$ is the transition probability function.
- Similarly for the DTMC process, $\{\tilde{\sigma}_m\}_{m\in\mathbb{Z}_+}$, induced from the parameter vector $\theta + \epsilon$.

Relative entropy in path-space

- Suggest performing parameter sensitivity analysis by comparing the path distributions utilizing relative entropy.
- Definition: The path-wise relative entropy of $Q^{\theta}_{0,M}$ w.r.t. $Q^{\theta+\epsilon}_{0,M}$ is

$$\mathcal{R}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) := \int \log\left(\frac{dQ^{\theta}_{0,M}}{dQ^{\theta+\epsilon}_{0,M}}\right) dQ^{\theta}_{0,M}$$

- Properties: (i) $\mathcal{R}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) \geq 0$ and
  (ii) $\mathcal{R}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) = 0$ iff $Q^{\theta}_{0,M} = Q^{\theta+\epsilon}_{0,M}$ a.e.
  (iii) $\mathcal{R}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) \leq \mathcal{R}\left(Q^{\theta}_{0,M+1} \,|\, Q^{\theta+\epsilon}_{0,M+1}\right)$

Relative entropy decomposition

- Its not difficult to see that path-wise relative entropy at stationary regime is decomposed into two parts; one independent of time and one that grows linearly with time:

$$\begin{aligned}\mathcal{R}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) &= \int_E \cdots \int_E \mu^{\theta}(\sigma_0) \prod_{i=0}^{M-1} p^{\theta}(\sigma_i,\sigma_{i+1}) \\ &\times \log \frac{\mu^{\theta}(\sigma_0)\prod_{i=0}^{M-1} p^{\theta}(\sigma_i,\sigma_{i+1})}{\mu^{\theta+\epsilon}(\sigma_0)\prod_{i=0}^{M-1} p^{\theta+\epsilon}(\sigma_i,\sigma_{i+1})} d\sigma_0 \cdots d\sigma_M \\ &= M\mathcal{H}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) + \mathcal{R}\left(\mu^{\theta} \,|\, \mu^{\theta+\epsilon}\right)\end{aligned}$$

where

- $\mathcal{H}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right)$ is the relative entropy rate (RER):

$$\mathcal{H}\left(Q^{\theta}_{0,M} \,|\, Q^{\theta+\epsilon}_{0,M}\right) = \mathbb{E}_{\mu^{\theta}}\left[\int p^{\theta}(\sigma,\sigma') \log \frac{p^{\theta}(\sigma,\sigma')}{p^{\theta+\epsilon}(\sigma,\sigma')} d\,\sigma'\right]$$

SA - Relative entropy rate

- For long times, RER is a sensible measure of parameter sensitivity.
- Properties:
  - Infer information about the path distribution. Consequently, it contains information not only for the invariant measure but also for the stationary dynamics such as metastable dynamics, exit times, time correlations, etc..
  - No need for explicit knowledge of invariant measure. Thus, it is suitable for reaction networks and non-equilibrium steady state systems.
  - RER is an observable of known test function $\Rightarrow$ tractable and statistical estimators can provide easily and efficiently its value.
  - Different ϵ's $\Rightarrow$ SA at different directions. However, the perturbed process is not needed to be simulated.

SA - Fisher information matrix

- RER reminds the discrete derivative SA method.
- Q: Can we avoid the (discrete) perturbation of θ by ϵ?
  A: Yes, by constructing the corresponding Fisher Information Matrix (FIM).
- It holds, under smoothness assumption on θ, that

$$\mathcal{H}\left(Q^{\theta}_{0,M} \mid Q^{\theta+\epsilon}_{0,M}\right) = \frac{1}{2}\epsilon^T \mathbf{F}_{\mathcal{H}}\big(Q^{\theta}_{0,M}\big)\epsilon + O(|\epsilon|^3)$$

  - where Fisher information matrix is defined as

$$\mathbf{F}_{\mathcal{H}}\big(Q^{\theta}_{0,M}\big) = \mathbb{E}_{\mu^{\theta}}\left[\int_E p^{\theta}(\sigma,\sigma')\nabla_{\theta}\log p^{\theta}(\sigma,\sigma')\nabla_{\theta}\log p^{\theta}(\sigma,\sigma')^T d\,\sigma'\right]$$

SA - Fisher information matrix

- FIM constitutes a derivative-free sensitivity analysis method.
- In optimal experimental design, the maximization of the determinant of the FIM constitutes the D-optimality test while the minimization of the trace of the inverse of the FIM constitutes the A-optimality test.
- Robustness on parameter perturbations as well as parameter identifiability can be inferred from the FIM.
  - Wider FIM implies a more robust model.
  - Steeper FIM implies more identifiable parameters.
- Eigenvalue analysis of FIM gives the most/least sensitive directions.

Preliminaries

- A continuous-time jump Markov process is fully determined by the transition rates, $c(\sigma, \sigma')$.
  - The total rate is defined by $\lambda(\sigma) = \sum_{\sigma'} c(\sigma, \sigma')$.
- For the parameter vector $\theta \in \mathbb{R}^K$, a CTMC, $\{\sigma_t\}_{t\in\mathbb{R}_+}$, is defined. Its path distribution in the interval $[0, T]$ is given by $Q^{\theta}_{0,T}\big(\{\sigma_t\}_{t\in[0,T]}\big)$.
- Similarly for the CTMC, $\{\tilde{\sigma}_m\}_{m\in\mathbb{Z}_+}$, induced from the parameter vector $\theta + \epsilon$.

Relative entropy decomposition

- Q: How to compute the Radon-Nikodym derivative, $\frac{dQ^{\theta}_{0,T}}{dQ^{\theta+\epsilon}_{0,T}}$?

  A: Apply Girsanov theorem which asserts that

$$\frac{dQ^{\theta}_{0,T}}{dQ^{\theta+\epsilon}_{0,T}}(\{\sigma_t\}) = \frac{\mu^{\theta}(\sigma_0)}{\mu^{\theta+\epsilon}(\sigma_0)} \exp\left\{\int_0^T \log \frac{c^{\theta}(\sigma_{s-},\sigma_s)}{c^{\theta+\epsilon}(\sigma_{s-},\sigma_s)} dN_s - \int_0^T [\lambda^{\theta}(\sigma_s) - \lambda^{\theta+\epsilon}(\sigma_s)]\, ds\right\}$$

- Substituting the above Girsanov formula to the path-wise relative entropy and exploiting the fact that we are at the steady state regime, we obtain that

$$\mathcal{R}\left(Q^{\theta}_{0,T} \mid Q^{\theta+\epsilon}_{0,T}\right) = T\mathcal{H}\left(Q^{\theta}_{0,T} \mid Q^{\theta+\epsilon}_{0,T}\right) + \mathcal{R}\left(\mu^{\theta} \mid \mu^{\theta+\epsilon}\right)$$

  - Relative Entropy Rate for CTMC is given by

$$\mathcal{H}\left(Q^{\theta}_{0,T} \mid Q^{\theta+\epsilon}_{0,T}\right) = \mathbb{E}_{\mu^{\theta}}\Big[\sum_{\sigma'\in E} c^{\theta}(\sigma,\sigma') \log \frac{c^{\theta}(\sigma,\sigma')}{c^{\theta+\epsilon}(\sigma,\sigma')} - (\lambda^{\theta}(\sigma) - \lambda^{\theta+\epsilon}(\sigma))\Big]$$

Fisher information matrix

- Generalize the notion of Fisher information theory to the case of CTMC processes.
- FIM is defined as

$$\mathbf{F}_{\mathcal{H}}(Q^\theta_{0,T}) := \mathbb{E}_{\mu^\theta}\left[\sum_{\sigma'\in E} c^\theta(\sigma,\sigma')\nabla_\theta \log c^\theta(\sigma,\sigma')\nabla_\theta \log c^\theta(\sigma,\sigma')^T\right]$$

Fisher information matrix

- Generalize the notion of Fisher information theory to the case of CTMC processes.
- FIM is defined as

$$\mathbf{F}_{\mathcal{H}}(Q_{0,T}^{\theta}) := \mathbb{E}_{\mu^{\theta}}\left[\sum_{\sigma' \in E} c^{\theta}(\sigma,\sigma')\nabla_{\theta}\log c^{\theta}(\sigma,\sigma')\nabla_{\theta}\log c^{\theta}(\sigma,\sigma')^{T}\right]$$

- Similar derivations can be obtained for time-inhomogeneous Markov chains, semi-Markov processes and probably for processes with memory.
- Sensitivity analysis on the logarithmic scale is also possible.

Connection with previous methods

- Pinsker (or Csiszar-Kullback-Pinsker) inequality:

$$||Q^{\theta}_{0,T} - Q^{\theta+\epsilon}_{0,T}||_{\mathrm{TV}} \le \sqrt{2\mathcal{R}\left(Q^{\theta}_{0,T} \mid Q^{\theta+\epsilon}_{0,T}\right)}$$

- If $Q^{\theta}_{0,T} = q^{\theta}_{0,T}dx$ and respectively $Q^{\theta+\epsilon}_{0,T} = q^{\theta+\epsilon}_{0,T}dx$,

$$\begin{aligned} |\mathbb{E}_{Q^{\theta}_{0,T}}[f] - \mathbb{E}_{Q^{\theta+\epsilon}_{0,T}}[f]| &\le ||f||_{\infty}||q^{\theta}_{0,T} - q^{\theta+\epsilon}_{0,T}||_1 \\ &\le ||f||_{\infty}\sqrt{2\mathcal{R}\left(q^{\theta}_{0,T} \mid q^{\theta+\epsilon}_{0,T}\right)} \end{aligned}$$

- Thus, if RER is insensitive to a parameter then any observable is insensitive to this particular parameter.
- Overall, in terms of specific observables, RER and FIM are conservative estimates of their parameter sensitivity.

Outline

Well-mixed reaction systems - p53

- The p53 gene model is an extensively studied system which plays a crucial role for effective tumor suppression in human beings as its universal inactivation in cancer cells suggests.
  - The p53 gene is activated in response to DNA damage and it constitutes a negative feedback loop with the oncogene protein Mdm2.

Event	Reaction	Rate
1	$\emptyset \to x$	$c_1(\sigma) = b_x$
2	$x \to \emptyset$	$c_2(\sigma) = a_x x + \frac{a_k y}{x+k} x$
3	$x \to x + y_0$	$c_3(\sigma) = b_y x$
4	$y_0 \to y$	$c_4(\sigma) = a_0 y_0$
5	$y \to \emptyset$	$c_5(\sigma) = a_y y$

- x corresponding to p53, y_0 to Mdm2-precursor while y corresponds to Mdm2. Parameter vector: $\theta = [b_x, a_x, a_k, k, b_y, a_0, a_y]^T$.

p53 - Concentrations

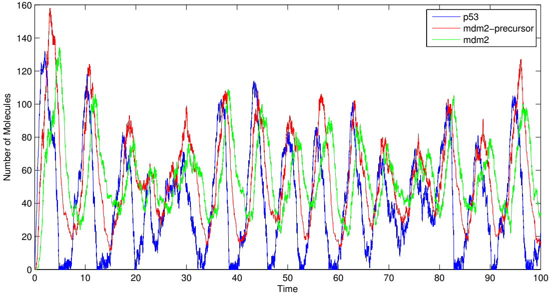


Figure: Molecule concentration of p53, Mdm2-precursor and Mdm2. Concentration oscillations as well as time delays (phase shifts) between the species are present due to the negative feedback loop. Furthermore, the concentration of p53 periodically hits the zero and since negative concentrations are not allowed, the stochastic characteristics of p53 are far from Gaussian.

p53 - RER

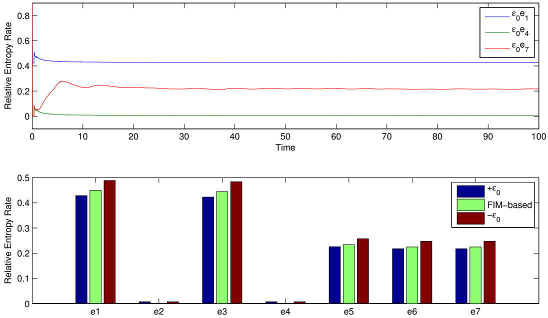


Figure: Upper panel: The Relative Entropy Rate in time for the parameter perturbation of b_x (blue), k (green) and a_y (red) by $+10\%$. The relaxation time of the RER as an observable is ultra fast. Lower panel: RER for various perturbation directions computed either directly (blue and red bars) or based on FIM (green bars).

p53 - FIM

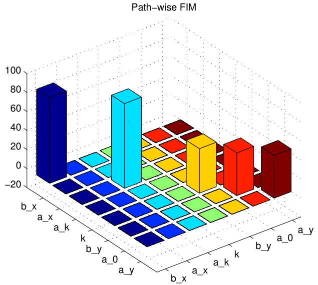


Figure: The proposed path-wise Fisher Information Matrix (left) based on RER as well as the FIM based on LNA (Komorowski et al., *Sensitivity, robustness, and identifiability in stochastic chemical kinetics models*, PNAS, 2011).

p53 - Comparison

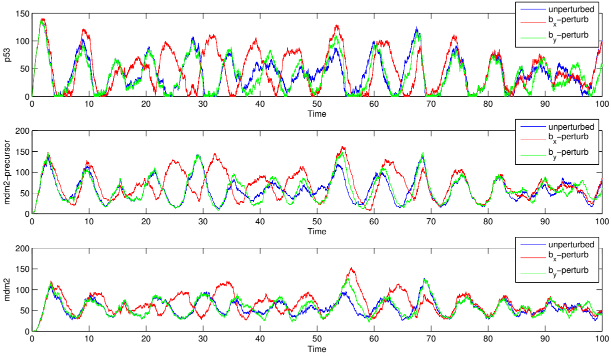


Figure: Time-series of the species for the unperturbed parameter regime (blue), when b_x is perturbed by +10% (red) as well as when b_y is perturbed by +10%. The same sequence of random numbers is used in all cases.

Langevin out-of-equilibrium process - Definition

- SDE system with N particles:

$$
\begin{aligned}
dq_t &= \frac{1}{m} p_t dt \\
dp_t &= -\mathbf{F}(q_t)dt - \frac{\gamma}{m} p_t dt + \sigma dB_t
\end{aligned}
$$

- Force: $\mathbf{F}(q) = \nabla_q V(q) + \alpha G(q)$
  - Interaction potential: $V(q) = \sum_{i,j<i} V_M(|q_i - q_j|)$
  - Morse potential: $V_M(r) = D_e(1 - e^{-a(r-r_e)})^2$
  - The divergence-free component ($\nabla_q \cdot G = 0$) is taken to be a simple antisymmetric force: $G_i(q) = q_{i+1} - q_{i-1}\ , \quad i = 1, ..., N$
- If $\alpha = 0$ then Langevin process is reversible. If $\alpha \neq 0$ then Langevin process is out of equilibrium.

Langevin out-of-equilibrium process - Scheme

- Sensitivity analysis on Morse parameters: $\theta = [D_e, a, r_e]$.
- Langevin process is time-discretized utilizing BBK integrator:

$$
\begin{aligned}
p_{i+\frac{1}{2}} &= p_i - \mathbf{F}(q_i)\frac{\Delta t}{2} - \frac{\gamma}{m} p_i \frac{\Delta t}{2} + \sigma \Delta W_i \\
q_{i+1} &= q_i + \frac{1}{m} p_{i+\frac{1}{2}} \Delta t \\
p_{i+1} &= p_{i+\frac{1}{2}} - \mathbf{F}(q_{i+1})\frac{\Delta t}{2} - \frac{\gamma}{m} p_{i+1} \frac{\Delta t}{2} + \sigma \Delta W_{i+\frac{1}{2}}
\end{aligned}
$$

- Transition probability of the discretized process:

$$P(q, p, q', p') = P(q'|q, p) P(p'|q', q, p)$$

where

- $P(q'|q,p) = \frac{1}{Z_0} e^{-\frac{m^2}{\sigma^2 \Delta t^3}\left|q' - q + \left(p - \mathbf{F}(q)\frac{\Delta t}{2} + p\frac{\Delta t \gamma}{2m}\right)\frac{\Delta t}{m}\right|^2}$
- $P(p'|q',q,p) = \frac{1}{Z_1} e^{-\frac{1}{\sigma^2 \Delta t}\left|(1 + \frac{\gamma \Delta t}{2m})p' - \left(\frac{m}{\Delta t}(q' - q) - \mathbf{F}(q')\frac{\Delta t}{2}\right)\right|^2}$

Langevin out-of-equilibrium process - FIM

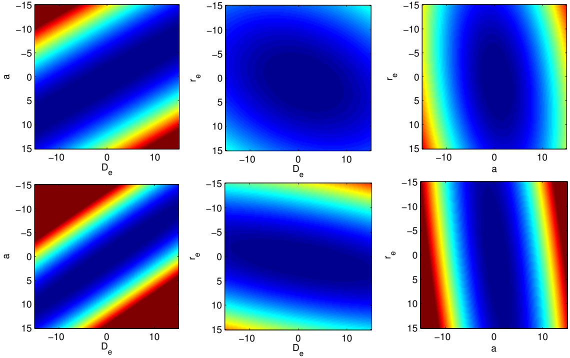


Figure: Upper plots: Level sets (or neutral spaces) for the reversible case ($\alpha = 0$). Lower plots: Level sets for the irreversible case ($\alpha = 0.1$).

ZGB - Definition

- ZGB (Ziff-Gulari-Barshad) is a simplified spatio-temporal CO oxidization model without diffusion.
- Despite being an idealized model, the ZGB model incorporates the basic mechanisms for the dynamics of adsorbate species during CO oxidation on catalytic surfaces.

Event	Reaction	Rate
1	$\emptyset \to CO$	$(1-\sigma(j)^2)k_1$
2	$\emptyset \to O_2$	$(1-\sigma(j)^2)(1-k_1)\frac{\#\text{vacant n.n.}}{\text{total n.n.}}$
3	$CO + O \to CO_2 + \text{des.}$	$\frac{1}{2}\sigma(j)(1+\sigma(j))k_2\frac{\#O\text{ n.n.}}{\text{total n.n.}}$
4	$O + CO \to CO_2 + \text{des.}$	$\frac{1}{2}\sigma(j)(\sigma(j)-1)k_2\frac{\#CO\text{ n.n.}}{\text{total n.n.}}$

Table: The rate of the kth event of the jth site given that the current configuration is σ is denoted by $c_k(j;\sigma)$ where n.n. stands for nearest neighbors.

ZGB - RER

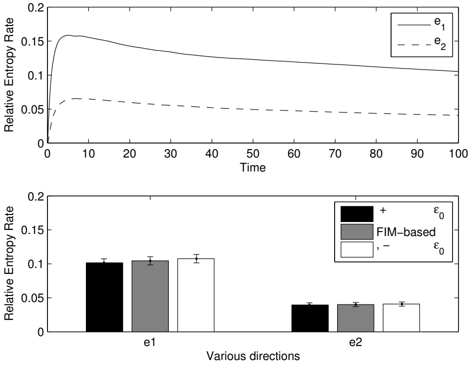


Figure: Upper plot: Relative entropy rate as a function of time for perturbations of both k_1 (solid line) and of k_2 (dashed line). An equilibration time until the process reach its metastable regime is evident. Lower plot: RER for various directions. The most sensitive parameter is k_1.

ZGB - Configurations

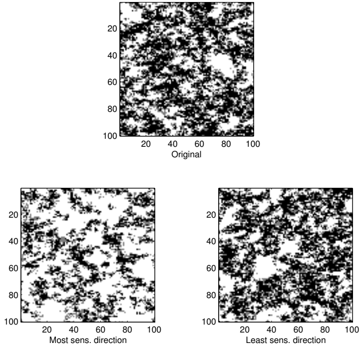


Figure: Typical configurations obtained by ϵ_0-perturbations of the most and least sensitive parameters. The comparison with the reference configuration reveals the differences between the most and least sensitive perturbation parameters.

Conclusions

- Sensitivity analysis of steady state stochastic processes utilizing RER and FIM.
  - Even though path-wise distributions was considered, the resulted observables where easy to compute.
  - Long-time stochastic dynamics are also taken into consideration.
  - No need to know the stationary distribution.
- *A Relative Entropy Rate Method for Path Space Sensitivity Analysis of Stationary Complex Stochastic Dynamics*, M.K. - Y.P., J. of Chemical Physics (2013).