Implementation of Nonreversible Metropolis-Hastings algorithms

Jin Hua WANG, Bo YUAN
Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200000, China
boyuan@cs.sjtu.edu.cn

Abstract. Metropolis-Hastings algorithm which keeps the detailed balance is the basic element for most Markov chain Monte Carlo sampling algorithms in which undermined Markov processes are reversible. Previous research shows that nonreversible Markov processes have a faster rate of convergence than reversible ones. Taking advantage of the “lifting” idea, this paper develops a general framework for designing Metropolis-Hastings algorithms breaking detailed balance and implements two new nonreversible Metropolis-Hastings algorithms based on the Gaussian proposal conditional probability and Langevin dynamics in the zero-mass limit respectively. Numerical simulations in one and two dimensions demonstrate that new nonreversible Metropolis-Hastings algorithms can speed up the convergence to target stationary distributions, which supports the theoretical finding and the design of our new algorithms.

1. Introduction
Markov Chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution as its stationary distribution. MCMC methods are primarily used for calculating numerical approximations of multi-dimensional integrals and parameter estimations in Bayesian statistical learning, statistical physics, computational biology and linguistics. Random walk Monte Carlo methods form its core, which includes Metropolis-Hastings algorithm [1,2], Gibbs sampling [3,4], Slice sampling [5], Multiple-try Metropolis [6] and Reversible-jump [7]. The Metropolis-Hastings algorithm is the basis of many related methods and considered as one of the most successful algorithms in the 20th century. It is, therefore, a significant work to improve the convergent speed of the Metropolis-Hastings algorithm.

The Metropolis-Hastings algorithm ensures the constructed Markov chain reversible and the generated samples drawn from its desired distribution by fulfilling the detailed balance condition (DBC). It is known that the DBC is a sufficient, unnecessary condition. Previous work [8~13] on the convergence of stochastic processes shows that nonreversible Markov processes have a rate of faster convergence than reversible ones. For the case of discrete state spaces, some practical samplers [14~16] exist, including the sampling approach via “lifting” state spaces [17~19]. For the case of continuous state spaces, relative study is rare or no such running algorithm found in practice [20,21].

This paper studies the method for implementing nonreversible Metropolis-Hastings algorithms in continuous state spaces as follows. In section 2, we first review the classical Metropolis-Hastings algorithm, then introduce the “lifting” idea into continuous state spaces to construct nonreversible Markov chains and propose an algorithmic framework of general nonreversible Metropolis-Hastings algorithms.
algorithms. We next present two different practical nonreversible Metropolis-Hastings algorithms in detail in section 3. Two numerical examples in section 4 are given to compare our new algorithms with the classical one. In section 5, we discuss the selection of parameters in our algorithms and possible further work.

2. Framework of General Nonreversible Metropolis-Hastings Algorithms through “Lifting” State Spaces

2.1. The Metropolis-Hastings Algorithm

The classical Metropolis-Hastings algorithm is described in the algorithm (MH) below.

\[ \text{MH: Metropolis-Hastings algorithm} \]

**Input:** Randomly select an initial value \( x_0 \) and set \( T \) a sufficient big number.

**Output:** Samples \( \{x_1, x_2, \cdots, x_T\} \) from the target distribution \( \pi \).

For \( (t = 1 \cdots T) \) do

1. Produce \( x' \propto f(x| x_{t-1}) = N(x_{t-1}, I) \) (the Gaussian distribution), and \( I \) is the identity matrix.
2. Calculate \( A(x'| x_{t-1}) = \min\{1, \frac{\pi(x') f(x_{t-1}| x')}{\pi(x_{t-1}) f(x'| x_{t-1})}\} \).
3. Produce a random number \( \alpha \propto U[0,1] \) (the uniform distribution with the value domain \([0,1]\)).
4. If \( \alpha \leq A(x'| x_{t-1}) \), then \( x_t = x' \), Otherwise \( x_t = x_{t-1} \).
End For.

This algorithm satisfies the DBC:

\[ \pi(x_t) f(x_{t+1}| x_t) A(x_{t+1}| x_t) = \pi(x_{t+1}) f(x_t| x_{t+1}) A(x_t| x_{t+1}) \]

and the Markov chain constructed is reversible. In the following two subsections, we build a nonreversible Markov process through “lifting” the target distribution state space and keeping the skew detailed balance, and present a framework of general nonreversible Metropolis-Hastings algorithms. We prove that any new Metropolis-Hastings algorithm within this framework breaks the DBC and keep the steady state distribution invariant.

2.2. “Lifting” in Continuous State Spaces and the Skew DBC

“Lifting” is a useful method to construct nonreversible Markov chains. Originally it is utilized in discrete state spaces. We apply it into continuous state spaces in the following to break the DBC.

Assume \( x \in \Omega \), \( x \propto \pi \) and \( P(x_i|x_j) \) is a transition probability from a state \( x_j \) to \( x_i \). As described above, \( P(x_i|x_j) = f(x_i|x_j) A(x_i|x_j) \) in the classical Metropolis-Hastings algorithm.

After introducing an auxiliary variable \( \xi \in \{-1, 1\} \), we have a larger state space \( \tilde{\Omega} = \Omega \times \{-1, 1\} \), which is twice biger than the original one. In this new space \( \tilde{\Omega} \), a state is represented by \( (x, \xi) \in \tilde{\Omega}, x \in \Omega, \xi \in \{-1, 1\} \). According to a certain local rule, inside the subspaces \( \Omega \times \{-1\} \) and \( \Omega \times \{1\} \), we introduce a mechanism of decomposition for a transition probability between any two states to satisfy:

\[ P(x_i|x_j) = \tilde{P}(x_i, \xi = 1|x_j, \xi = 1) + \tilde{P}(x_i, \xi = -1|x_j, \xi = -1) \]

where \( \tilde{P}(x_i, \xi = 1|x_j, \xi = 1) > 0, \tilde{P}(x_i, \xi = -1|x_j, \xi = -1) > 0 \).

We require that the following skew DBC holds, i.e. for any \( i \neq j \),

\[ P(x_i, \xi = 1|x_j, \xi = 1) = \tilde{P}(x_j, \xi = -1|x_i, \xi = -1) \tilde{\pi}(x_i, \xi = 1) \]

\[ \tilde{\pi} = \frac{1}{2} (\pi, \pi) \]

and a transition probability between the two subspaces \( \Omega \times \{-1\} \) and \( \Omega \times \{1\} \) satisfies:

\[ \tilde{P}(x_j, -\xi|x_i, \xi) \geq 0 \]

To make the transition probability equation above hold, we set

\[ \tilde{P}(x, -\xi|x, \xi) = 1 - \int_{\Omega} \tilde{P}(y, -\xi|x, \xi) \, dy \]

and
\[ \bar{P}(y, -\xi|x, \xi) = \delta_{x,y} \bar{P}(y, -\xi|x, \xi). \]  

(4)

Now let us approve a Markov chain which comply with the requirement like the skew DBC (2) and setting above such as the equations (3) and (4) has the distribution \( \bar{\pi} \) as its invariant probability distribution. Assume \( A \) a domain in \( \Omega \),

\[
\int_{\Omega} \bar{P}(A|x, \xi) \bar{\pi}(x, \xi) dx = \bar{\pi}(A, \xi) = \bar{\pi}(A, -\xi)
\]

The equation \( \int_{\Omega} \bar{P}(A|x, \xi) \bar{\pi}(x, \xi) dx \) is the definition of an invariant distribution of a Markov process.

2.3. Framework of General Nonreversible Metropolis-Hastings Algorithms

According to the “lifting” idea and requirements discussed above, we, in this subsection, propose a new framework for constructing general nonreversible Metropolis-Hastings algorithms in which both the decomposition equation of the transition probability (1) and the skew DBC (2a) hold, So any nonreversible Metropolis-Hastings algorithm under this framework converges to the desired distribution.

Assume \( f(x_j|x_i) \) be the proposal conditional probability from a state \( x_j \) to \( x_i \) in the original space in the classical Metropolis-Hastings algorithm, \( \pi \) the original target distribution.

If \( f(x_j|x_i) \) has the following decomposition:

\[
f(x_j|x_i) = f(x_j, \xi|x_i, \xi) + f(x_j, -\xi|x_i, -\xi), \quad f(x_j, \xi|x_i, \xi) \neq f(x_j, -\xi|x_i, -\xi)
\]

and \( \bar{P}(x_j, \xi|x_i, \xi) \) satisfies

\[
\bar{P}(x_j, \xi|x_i, \xi) = \min \left\{ \frac{f(x_j, -\xi|x_i, -\xi) \bar{\pi}(x_i, -\xi)}{f(x_j, \xi|x_i, \xi) \bar{\pi}(x_i, \xi)}, \frac{f(x_j, \xi|x_i, \xi) \bar{\pi}(x_i, \xi)}{f(x_j, -\xi|x_i, -\xi) \bar{\pi}(x_i, -\xi)} \right\} f(x_i, \xi|x_j, \xi),
\]

then, for any \( i \neq j \)

\[
\bar{P}(x_i, \xi = 1|x_j, \xi = 1) + \bar{P}(x_i, \xi = -1|x_j, \xi = -1) = \min \left\{ \frac{f(x_j, \xi = -1|x_i, \xi = -1) \bar{\pi}(x_i, \xi = -1)}{f(x_j, \xi = 1|x_i, \xi = 1) \bar{\pi}(x_i, \xi = 1)}, \frac{f(x_j, \xi = 1|x_i, \xi = 1) \bar{\pi}(x_i, \xi = 1)}{f(x_j, \xi = -1|x_i, \xi = -1) \bar{\pi}(x_i, \xi = -1)} \right\} f(x_i, \xi = 1|x_j, \xi = 1) \]

\[
= \min \left\{ \frac{f(x_j, \xi = 1|x_i, \xi = 1) \bar{\pi}(x_i, \xi = 1)}{f(x_j, \xi = -1|x_i, \xi = -1) \bar{\pi}(x_i, \xi = -1)} \right\} f(x_i, \xi = 1|x_j, \xi = 1) = f(x_i, \xi = 1|x_j, \xi = 1).
\]
\[ P(x_j|x_i) = P(x_i|x_j). \]

The Equation (1) holds.

Meanwhile,

\[
\begin{align*}
\bar{P}(x_j, \xi = 1 | x_i, \xi = 1) &= \min \left\{ 1, \frac{f(x_j, \xi = 1 | x_i, \xi = 1)\bar{p}(x_j, \xi = 1)}{f(x_i, \xi = 1 | x_j, \xi = 1)\bar{p}(x_i, \xi = 1)} \right\} f(x_i, \xi = 1 | x_j, \xi = 1) \\
&= \min \left\{ 1, \frac{f(x_j, \xi = 1 | x_i, \xi = 1)\bar{p}(x_j, \xi = 1)}{f(x_i, \xi = 1 | x_j, \xi = 1)\bar{p}(x_i, \xi = 1)} \right\} f(x_i, \xi = 1 | x_j, \xi = 1) \\
&= \bar{P}(x_j, \xi = 1 | x_i, \xi = 1) \bar{p}(x_i, \xi = 1) \\
\end{align*}
\]

The Equation (2a) holds.

It is concluded that a Metropolis-Hastings algorithm in which its proposal conditional probability satisfies the decomposition like the equation (5), its transition probability selected as (6) and (3) as well as (4) constructs a nonreversible Markov chain and produces samples from the extended target distribution \( \bar{p} \). We, therefore develop a general framework for designing nonreversible Metropolis-Hastings algorithms which is described in the algorithm (GNMH) as follows.

**GNMH:** a general framework of nonreversible Metropolis-Hastings algorithms

**Input:** Randomly select initial value \((x_0, \xi_0)\) and set \( T \) a sufficient big number.

**Output:** Samples \( \{x_1, x_2, \ldots, x_T\} \) from the target distribution \( \pi \).

For \( (t = 1 \ldots T) \) do

1. Produce \((x', \xi_{t-1}) \propto f(\cdot, \xi_{t-1} | x_{t-1}, \xi_{t-1})\).

2. Calculate \( A(x', \xi_{t-1} | x_{t-1}, \xi_{t-1}) = \min \{ 1, \frac{f(x_{t-1}, -\xi_{t-1} | x', -\xi_{t-1})\bar{p}(x', -\xi_{t-1})}{f(x_{t-1}, \xi_{t-1} | x', \xi_{t-1})\bar{p}(x_{t-1}, \xi_{t-1})} \} \).

3. Produce a random number \( \alpha \propto U[0,1] \) (the uniform distribution with the value domain \([0,1]\)).

4. If \( \alpha \leq A(x', \xi_{t-1} | x_{t-1}, \xi_{t-1}) \), then \((x_t, \xi_t) = (x', \xi_{t-1})\), Otherwise \((x_t, \xi_t) = (x_{t-1}, -\xi_{t-1})\).

End For.

3. Two Implementations

In this section, we first present a simple method to realize the decomposition of the proposal conditional probability. Next based on the common normal distribution and Langevin dynamics in the zero-mass limit [22] respectively, we implement two different nonreversible Metropolis-Hastings algorithms.

A simple decompose method to satisfy the equation (5) is

\[
\begin{align*}
f(x_i, \xi | x_j, \xi) &= \frac{1}{2} (f(x_i | x_j) + g(x_i | x_j)), \\
f(x_i, -\xi | x_j, -\xi) &= \frac{1}{2} (f(x_i | x_j) - g(x_i | x_j))
\end{align*}
\]

Here \( g(x_i | x_j) \neq 0 \) is one of any determine functions.

It is direct to derive that the equations (6) and (7) make the equations (1) and (2a) hold, but the DBC broken. We give two selections for \( f(x_i | x_j) \) and \( g(x_i | x_j) \) in the equation (7) which result in two practical nonreversible Metropolis-Hastings algorithms as follows.

3.1. \( f(x_i | x_j) = \mathcal{N}(x_j, I) \) and \( g(x_i | x_j) \) is a nonzero constant vector \( \alpha \)

When \( f(x_i | x_j) \) takes the most commonly used normal distribution \( \mathcal{N}(x_j, I) \) (\( I \) is the identity matrix) and \( g(x_i | x_j) \) a nonzero constant vector \( \alpha \), we achieve the algorithm (NMH1) below.

**NMH1:** a nonreversible Metropolis-Hastings algorithm based on the normal distribution

**Input:** Randomly select an initial value \((x_0, \xi_0)\) and set \( T \) a sufficient big number as well as \( \alpha \).

**Output:** Samples \( \{x_1, x_2, \ldots, x_T\} \) from the target distribution \( \pi \).

For \( (t = 1 \ldots T) \) do

1. If \( \xi_{t-1} = 1 \) then
Produce $x' \propto \mathcal{N}(x_t + \alpha, I)$, calculate $A(x'|x_{t-1}) = \min(1, \frac{\pi(x')}{\pi(x_{t-1})})$.

Else
Produce $x' \propto \mathcal{N}(x_t - \alpha, I)$, calculate $A(x'|x_{t-1}) = \min(1, \frac{\pi(x')}{\pi(x_{t-1})})$.

Endif

2. Produce a random number $\beta \propto U[0,1]$ (the uniform distribution with the value domain [0,1]).

3. If $\beta \leq A(x'|x_{t-1})$, then $(x_t, \xi_t) = (x', \xi_{t-1})$, Otherwise $(x_t, \xi_t) = (x_{t-1}, -\xi_{t-1})$.

End For.

3.2 Langevin dynamics in the zero-mass limit
Considering the Langevin dynamics (stochastic differential equation) in the zero-mass limit, we select
\[ f(x'|x_{t-1}) = \mathcal{N}(x_{t-1} - \frac{1}{D(x_{t-1})} \nabla_x U(x)|_{x=x_{t-1}}) dt + \phi(x_{t-1}) dt, \]
\[ g(x'|x_{t-1}) = -\frac{\nabla_x U(x)|_{x=x_{t-1}}}{\sqrt{D(x_{t-1})}} + \phi(x_{t-1}) dt, \]
where $\phi_i(x_{t-1}) = \sum_{j=1}^{m} \frac{\partial q_i(x)}{\partial x_j}|_{x=x_{t-1}}$, $\phi_i(x_{t-1}) = \sum_{j=1}^{m} \frac{\partial q_j(x)}{\partial x_i}|_{x=x_{t-1}}$, $U(x) = -\ln \pi(x)$, $D(x)$ is a positive semidefinite symmetric matrix and $Q(x)$ an anti-symmetric matrix. Our choice leads to the following algorithm (NMH2).

NMH2: a nonreversible Metropolis-Hastings algorithm based on Langevin dynamics in the zero-mass limit
Input: Randomly select an initial value $(x_0, \xi_0)$ and set $T$ a sufficient big number.
Output: Samples $\{x_1, x_2, \ldots, x_T\}$ from the target distribution $\pi$.

For $(t = 1 \ldots T)$ do
1. If $\xi_{t-1} = 1$ then
Produce $x' = x_{t-1} - \frac{1}{\sqrt{D(x_{t-1})}} \nabla_x U(x)|_{x=x_{t-1}} dt + (\phi(x_{t-1}) + \phi(x_{t-1})) dt + \sqrt{D(x_{t-1})} dw(t)$, calculate $A(x'|x_{t-1}) = \min(1, \frac{f(x_{t-1}|x')}{-f(x_{t-1}) + f(x_{t-1}|x')})$.

Else
Produce $x' = x_{t-1} - \frac{1}{\sqrt{D(x_{t-1})}} \nabla_x U(x)|_{x=x_{t-1}} dt + (\phi(x_{t-1}) - \phi(x_{t-1})) dt + \sqrt{D(x_{t-1})} dw(t)$, calculate $A(x'|x_{t-1}) = \min(1, \frac{f(x_{t-1}|x')}{-f(x_{t-1}) + f(x_{t-1}|x')})$.

Endif

2. Produce a random number $\alpha \propto U[0,1]$ (the uniform distribution with the value domain [0,1]).

3. If $\alpha \leq A(x'|x_{t-1})$, then $(x_t, \xi_t) = (x', \xi_{t-1})$, Otherwise $(x_t, \xi_t) = (x_{t-1}, -\xi_{t-1})$.

End For.

4. Numerical Examples

4.1 One Dimensional System with Two Stable States
In our first example, the probability distribution is $e^{-U(x)}$ and
\[ U(x) = \frac{1}{14}(x + 4)(x + 1)(x - 1)(x - 3) + \frac{1}{2}. \]

It has two meta-stable states.

We sample from it with MH and NMH1 with $\alpha = 0.5$ and $\alpha = 0.5$ respectively. In all algorithms, $T = 10^5$ the first $1/2T$ is a burn-in stage and samples are collected in the last half $1/2T$. The results of two samplers are illustrated in the following three figures. In each of these three figures, the red line is drawn from the true target distribution and the blue line is drawn from the target distribution with different algorithms.
Figure 1 Sampling result from $e^{-U(x)}$ with MH ($T = 10^5$)

Figure 2 Sampling result from $e^{-U(x)}$ with NMH1 ($T = 10^5, \alpha = 0.5$)
4.2. Two-dimensional System with Two Stable States

In our second example, the probability distribution is \( e^{-U(x,y)} \) and

\[
U(x, y) = \frac{1}{4} (x^2 - 1)^2 + \frac{1}{2} y^2
\]

It has two meta-stable states \((-1,0)\) and \((1,0)\).

We sample from it with MH and NMH2 with \( D(x, y) = I \) (the identity matrix), \( Q(x, y) = \begin{pmatrix} 0 & -5 \\ 5 & 0 \end{pmatrix} \) respectively. In those algorithms, \( T = 10^5 \) and the first half \( 1/2T \) is a burn-in stage and samples are collected in the last half \( 1/2T \) as we did in our first example. The results of two samplers are illustrated in the following figure.

These figures clearly demonstrate that our new nonreversible Metropolis-Hastings algorithms accelerate the convergence to the target distribution than the classical reversible one while they have the same time complexity \( O(T) \).
5. Conclusion
This paper proposes a new framework for constructing general nonreversible Metropolis-Hastings algorithms and presents two implements based on the Gaussian proposal conditional probability and Langevin dynamics in the zero-mass limit respectively. Moreover, it is not hard to find that NMH1 is a special case of NMH2. Numerical examples illustrate that our new Metropolis-Hastings algorithms improve the rate of convergence of sampling.

Considering the local geometric curvature of the target distribution function, we may take the inverse of the expected or experience Fisher information matrix as the positive semidefinite symmetric matrix $D(\chi)$ in our algorithm NMH2. Furthermore, The Langevin dynamics in NMH2 can be applied in the Hamiltonian system to allow the anti-symmetric matrix $Q(\chi)$ more choices even with the Fisher information [23].

It is an unsolved and interesting task to determine the parameter vector $\alpha$ in our algorithm NMH1 so as to attain an optimal sampling speed. We will further work on this issue and other possible implementations within our general nonreversible Metropolis-Hastings algorithmic framework for more efficient sampling.

References
[1] Hastings W K. Monte Carlo sampling methods using Markov chains and their applications[J]. Biometrika., 1970, 57(1): 97-109.
[2] Chib S, Greenberg E. Understanding the metropolis-hastings algorithm[J]. The american statistician, 1995, 49(4): 327-335.
[3] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images[J]. IEEE Transactions on pattern analysis and machine intelligence, 1984, 6:721-741.
[4] Gilks W R, Wild P. Adaptive rejection sampling for Gibbs sampling[J]. Applied Statistics, 1992, 41(2):337-348.
[5] Neal R M. Slice sampling[J]. Annals of statistics, 2003, 31(3):705-767.
[6] Liu J S, Liang F, Wong W H. The multiple-try method and local optimization in Metropolis sampling[J]. Journal of the American Statistical Association, 2000, 95(449):121-34.
[7] Green P J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination[J]. Biometrika, 1995, 82(4):711-32.
[8] Hwang C R, Hwang-Ma S Y, Sheu S J. Accelerating diffusions[J]. The Annals of Applied Probability, 2005, 15(2):1433-44.
[9] Hwang C R, Hwang-Ma S Y, Sheu S J. Accelerating gaussian diffusions[J]. The Annals of Applied Probability, 1993, 3(3):897-913.
[10] Diaconis P, Holmes S, Neal R M. Analysis of a nonreversible Markov chain sampler[J]. Annals of Applied Probability, 2000, 10(3):726-752.
[11] Lelievre T, Nier F, Pavliotis G A. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion[J]. Journal of Statistical Physics, 2013, 152(2):237-74.
[12] Wu S J, Hwang C R, Chu M T. Attaining the optimal Gaussian diffusion acceleration[J]. Journal of Statistical Physics, 2014, 155(3):571-90.
[13] Duncan A B, Lelievre T, Pavliotis G A. Variance reduction using nonreversible Langevin samplers[J]. Journal of statistical physics, 2016, 163(3):457-91.
[14] Turitsyn K S, Chertkov M, Vucelja M. Irreversible Monte Carlo algorithms for efficient sampling[J]. Physica D: Nonlinear Phenomena, 2011, 240(4):410-414.
[15] Hukushima K, Sakai Y. An irreversible Markov-chain Monte Carlo method with skew detailed balance conditions[J]. Journal of Physics: Conference Series 2013, 473(1):012012, IOP Publishing.
[16] Ohzeki M, Ichiki A. Langevin dynamics neglecting detailed balance condition[J]. Physical Review E, 2015, 92(1):012105.
[17] Chen F, Lovász L, Pak I. Lifting Markov chains to speed up mixing[C]. In Proceedings of the thirty-first annual ACM symposium on Theory of computing, May 1, 1999, 275-281.
[18] Li W, Dai H. Accelerating distributed consensus via lifting Markov chains[C]. In Information Theory, 2007, ISIT 2007, 2881-2885.
[19] Vucelja M. Lifting—A nonreversible Markov chain Monte Carlo algorithm[J]. American Journal of Physics, 2016, 84(12):958-68.
[20] Bierkens J. Non-reversible metropolis-hastings[J]. Statistics and Computing, 2016, 26(6):1213-28.
[21] Bierkens J, Roberts G. A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model[J]. The Annals of Applied Probability, 2017, 27(2):846-82.
[22] Wang J, Yuan B. Curl flux induced drift in stochastic differential equations in the zero-mass limit[J]. Physica A: Statistical Mechanics and its Applications, 2016, 462:898-911.
[23] Ma Y, Chen T, Fox E. A complete recipe for stochastic gradient MCMC[C]. In Advances in Neural Information Processing Systems, 2015, 2917-2925.