Parametric Inverse Simulation

Zenna Tavares
MIT
zenna@mit.edu

Armando Solar Lezama
MIT
asolar@csail.mit.edu

Abstract

We introduce the theory of *parametric inversion*, which generalizes function inversion to non-invertible functions. We define parametric inverses for several common functions, and present an algorithm to symbolically transform a composition of functions (i.e., a program) into a corresponding parametric inverse. We then outline a variational approach to sampling from a conditional distribution as a form of constrained parametric inversion.

Recent advances in probabilistic programming [6, 5] and deep-generative modeling [4, 7], have dramatically expanded the kinds of probabilistic models we can express and learn. Unfortunately, Bayesian inference procedures have not kept pace; there is a growing chasm between our ability to simulate complex probability distribution, and invert them for probabilistic inference.

In such complex probabilistic models it is often difficult to infer latent variables that are *consistent* with the observed values, let alone drawn from the correct posterior distribution. For example, we can view vision as inference of probable scenes (geometry, lighting, material, etc) that render to an observed image, but forward simulation of a prior distribution will in practice generate scenes which are simply inconsistent with our observations.

We propose an alternative approach: to simulate a random variable in reverse, starting with the observations. Inferred latent values are then consistent by construction. To define inverse simulation when a random variable is not invertible, we present *parametric inversion*, which generalizes function inversion to non-injective functions. A non-injective function $f : X \to Y$ is not invertible because it lacks a unique right inverse, a function $f^{-1} : Y \to X$ which for all y satisfies:

$$f(f^{-1}(y)) = y \tag{1}$$

To generalize inversion, a parametric inverse represents a *set* of right-inverses parametrically, as a mapping from a parameter space to a function in a set.

Parametric inversion of random variables leads to a novel formulation of Bayesian inference by inverse simulation. In this abstract, we: (i) introduce the concept of parametric inversion, (ii) present an algorithm to construct a parametric inverse from a composition of these primitives, and (iii) outline its extension to random variables for conditional sampling.

1 Example: Bayesian Inference as Constrained Inversion

To demonstrate how Bayesian inference requires function inversion, consider the following model:

$$x \sim exponential(\lambda = 1) \qquad y \sim logistic(\mu = x, s = 1)$$

We can express this model as a pair of random variables (figure 1), i.e., transformations of $\omega \in [0, 1]$:

$$exponential(\omega; \lambda) = -\ln(1 - \omega)/\lambda \qquad x(\omega_1) = exponential(\omega_1, 1)$$
$$logistic(\omega; \mu, s) = \mu + s\ \ln(\omega/(1 - \omega)) \qquad y(\omega_1, \omega_2) = logistic(\omega_2, x(\omega_1), 1)$$

To conditionally sample a value of x which is consistent with observation $y = c$, let $(\omega_1^*, \omega_2^*) = y^{-1}(c)$, then evaluate $x(\omega_1^*)$. Hence, conditioning requires inverting y, but what if y is not invertible? How can we draw samples from x that are not only consistent, but from the posterior $P(x \mid y = c)$?

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

2 Parametric Inversion

Parametric inversion generalizes function inversion to non-invertible functions. A parametric inverse of a function $f : X \to Y$ is a parameterized function $f^{-1} : Y \times \Theta \to X$ which maps an element y to a single element of its preimage: $f(y; \theta) \in \{x \mid f(x) = y\}$. The parameter θ determines which element of the preimage is returned.

In contrast to a conventional inverse, a parametric inverse always exists. For example, while $f(x) = |x|$ is not invertible, we can define a parameter space $\theta \in \{-1, 1\}$ and parametric inverse $f^{-1}(y; \theta) = \theta \cdot y$, where θ determines whether the positive or negative inverse element is returned.

If a function takes multiple inputs, its parametric will return a tuple. For example, inverse multiplication (we denote $\times^{-1}$) maps a value z to two values whose product is z: $\times^{-1}(z; \theta) = (\theta, z/\theta)$, where $\theta \in \mathbb{R} \setminus 0$. Parametric inverses are not unique; it is equally valid to define $\times^{-1}(z; \theta) = (z/\theta, \theta)$.

Many functions map between a value and its preimage; we reserve the term parametric inverse for those which are sound and complete in this task.

Definition 1. $f^{-1} : Y \times \Theta \to X$ is a sound and complete parametric inverse of $f : X \to Y$ if $\forall y$:

$$\{f^{-1}(y; \theta) \mid \theta \in \Theta\} = \{x \mid f(x) = y\} \tag{2}$$

Condition 2 asserts that parametric inverses are (i) sound: for any θ, $f^{-1}(y; \theta)$ is an element of the preimage of y, and (ii) complete: there exists θ corresponding to every element of the preimage. Inverse addition as $+^{-1}(z; \theta_1, \theta_2) = (\theta_1, \theta_2)$ is complete but not sound; for any z, there exists a θ_1 and θ_2 which sum to z, but there are many more pairs which do not. $+^{-1}(z; \theta) = (0, z)$ is sound but not complete; there exists no θ such that for instance $+^{-1}(0; \theta) = (-2, 2)$, despite the fact that $-2 + 2 = 0$. Hence, neither of these are valid parametric inverses of addition.

A parametric inverse has a dual interpretation. If y is fixed, varying the parameter θ varies which element of the preimage of y is returned. If θ is fixed, f^{-1} is a function of only y and a right-inverse which satisfies equation 1. These two alternatives are useful for different applications.

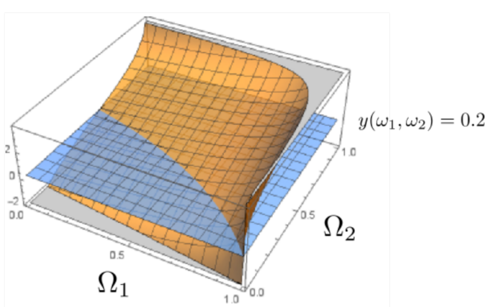

Figure 1: Bayesian parameter estimation. Random variable y is a function of sample space. Intersection of y with plane is subset of $\Omega_1 \times \Omega_2$ which is consistent with $y = 0.2$. Varying θ in parametric inverse $y^{-1}(0.2; \theta)$ covers this set

f	Θ	f^{-1}
$x + y$	$\mathbb{R}$	$(\theta, z - \theta)$
$x - y$	$\mathbb{R}$	$(z + \theta, \theta)$
$x \cdot y$	$\mathbb{R} \setminus 0$	$(z/\theta, \theta)$
x/y	$\mathbb{R}$	$(z \cdot \theta, \theta)$
x^y	$\mathbb{R}$	$(\theta, log_\theta(z))$
$log_x(y)$	$\mathbb{R} \setminus 0$	(θ, θ^z)
$min(x, y)$	$\mathbb{R}^+$	$(z, z + \theta)$
$max(x, y)$	$\mathbb{R}^+$	$(z, z - \theta)$
$sin(x)$	$\mathbb{Z}$	$asin(x) + 2\pi\theta$
$cos(x)$	$\mathbb{Z}$	$acos(x) + 2\pi\theta$

Figure 2: Primitive parametric inverses. Column f contains forward non-injective functions. Θ is the parameter space used for parametric inverses shown in column f^{-1}.

3 Parametric Inversion of Composite Functions

Arbitrarily complex functions can be found by composing functions from a small set of primitive functions. Given functions $f_2 : X \to Y$, $f_1 : Y \to Z$, and composition $f = f_1 \circ f_2$, to construct a parametric inverse $f^{-1} : Z \times \Theta \to X$ we substitute f_1 and f_2 with their corresponding parametric inverses f_1^{-1} and f_2^{-1}, and reverse the order of composition:

$$f^{-1}(z; \theta_1, \theta_2) = f_2^{-1}(f_1^{-1}(z, \theta_1), \theta_2) \tag{3}$$

For illustration, let $f = \sin \circ \cos$, then $f^{-1}(z) = \sin^{-1}(\cos^{-1}(z;\theta_1), \theta_2)$ with parameter space $\Theta = \Theta_1 \times \Theta_2$. In an extension of notation, we overload the composition operator $\circ$ to denote this inverse composition of parametric inverses as $\cos^{-1} \circ \sin^{-1}$.

By construction f^{-1} will be a parametric inverse of f on some subset of $\Theta_1 \times \Theta_2$. Continuing our example, for most θ, $\sin^{-1}(y;\theta)$ will return a value not within $[-1, 1]$, the domain of $\cos^{-1}$. f^{-1} is undefined for these parameter values, and hence a *partial*, not total parametric inverse.

Definition 2. If a function $f^{-1} : Y \times \Phi \to X$ does not satisfy condition 2, it is a partial parametric inverse if there exists a $\Theta \subset \Phi$ such that $f^{-1} : Y \times \Theta \to Y$ satisfies condition 2.

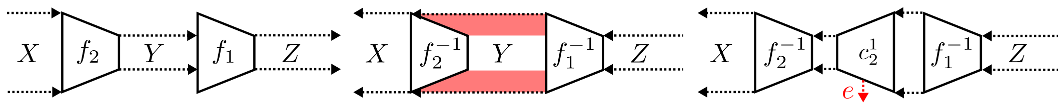

Figure 3: The forward function (left) is a total function. Its parametric inverse (center) is partial, because there exists parameter values which generate values (in red) on which f_2^{-1} is undefined. The solution used in approximate parametric inverses is to insert a restriction c between the two.

Parametric inversion extends to longer chains of composed functions. For convenience, if $f^{-1} = f_2^{-1} \circ f_1^{-1}$, then f^{-1} is defined as in equation 3,

Algorithm 1. Given $f = f_1 \circ f_2 \cdots \circ f_n$, substitute f_i with f_{n-i}^{-1} to construct $f^{-1} = f_n^{-1} \circ f_{n-1}^{-1} \circ \cdots \circ f_1^{-1}$, where f_i^{-1} is a parametric inverse of f_i. f^{-1} is a partial parametric inverse of f with a parameter space $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$, where Θ_i is the parameter space of primitive f_i^{-1}.

4 Approximate Parametric Inversion

If the undefined subset of the parameteric space is large or complex, as typically is the case, a parametric inverse will be useless in practice. We present approximations which force a parametric inverse to be total, but may be unsound.

Definition 3. An approximate parametric inverse is a function $\widetilde{f}^{-1} : Y \times \Theta \to X \times \mathbb{R}^+$, which returns an approximate inverse element $\tilde{x} \in X$ and a non-negative error term $e \in \mathbb{R}^+$. $\widetilde{f}^{-1}$ is an approximate parametric inverse of $f : X \to Y$ if $e = 0$ implies $\tilde{x} \in \{x \mid f(x) = y\}$.

To transform a partial parametric inverse into an approximate we must make it: (1) well defined on all of its parameter space, and (ii) output an error term of 0 when unsound.

To solve the first problem we insert a function c_2^1 between each pair of primitives f_{i-1}^{-1} (see figure 3) and f_i^{-1}, which takes any value in the domain of f_i^{-1} and restricts it to the domain of f_{i-1}^{-1}. For example, $\cos^{-1} \circ \sin^{-1}$ becomes $\cos^{-1} \circ clip_{-1,1} \circ \sin^{-1}$, where the $clip_{a,b}$ returns its input clamped to the closest point in the interval $[a, b]$, and the distance between its input and output as an error term:

$$clip_{a,b}(x) = (b, x - b) \text{ if } x \geq b \mid (a, a - x) \text{ if } x \leq a \mid (x, 0) \text{ otherwise}$$

The error term of the composition is then simply the sum of the error terms of its elements.

4.0.1 Approximate Composition Chain

Algorithm 2 extends this procedure to a chain of composed functions. We first define $\widetilde{f}_1^{-1} \circ \widetilde{f}_2^{-1}$:

Let $(\tilde{x}_1, e_1) = \widetilde{f}_1^{-1}(y;\theta_1)$ and $(\tilde{x}_2, e_2) = \widetilde{f}_2^{-1}(\tilde{x}_1; e_1)$, then $\widetilde{f}^{-1} = \widetilde{f}_1^{-1} \circ \widetilde{f}_2^{-1}$ is defined as:

$$\widetilde{f}^{-1}(y;\theta_1,\theta_2) = (\tilde{x}_2, e_1 + e_2) \tag{4}$$

Algorithm 2. Given a partial parametric inverse $f^{-1} = f_n^{-1} \circ f_{n-1}^{-1} \circ \cdots \circ f_1^{-1}$, substitute f_i^{-1} with $c_{i,j} \circ f_i^{-1}$ to construct $\widetilde{f}^{-1} = f_n^{-1} \circ c_n^{n-1} \circ f_{n-1}^{-1} \circ \cdots \circ f_2^{-1} \circ c_2^1 \circ f_1^{-1}$. $\widetilde{f}^{-1}$ is an approximate parametric inverse of f, where c_{i-1}^{i} restricts values in the codomain of f_i^{-1} to the domain of f_{i-1}^{-1}.

Often the model we wish to invert is not expressible in the form $f_i \circ f_{i+1} \circ \cdots \circ f_n$. However, the basic approach of algorithm 1 and 2 - to substitute each primitive operation with its parametric inverse and reverse the direction of information flow - extends to richer classes of computation.

5 Variational Inversion of Random Variables

We propose to extend parametric inversion to conditionally sample from a random variable. A random variable is a function $X : \Omega \to T$ where Ω is sample space equipped with a measure $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$. To sample from X conditioned on observation $X = c$ means to (i) construct a predicate $Y : \Omega \to \{0, 1\}$ where $Y(\omega) := X(\omega) = c$, (ii) sample $\omega \in A$ with probability $\mathbb{P}(\omega)/\mathbb{P}(A)$, where $A = \{\omega \mid \omega \in \Omega, Y(\omega) = 1\}$ is the conditioning event, and (iii) compute $X(\omega)$.

Step (ii) implicitly involves the inversion of Y; we can use algorithm 1 to explicitly transform $Y : \Omega \to \{0, 1\}$ into a parametric inverse $Y^{-1} : \{0, 1\} \times \Theta \to \Omega$. To simplify notation we ignore scenarios where our condition does not hold and denote $Y_1^{-1}(\theta) = Y^{-1}(1; \theta)$. Then for any $\theta \in \Theta$, $X(Y_1^{-1}(\theta))$ is a value which is *consistent* with Y but may well not be sampled from the correct distribution $P(X|Y)$. To sample correctly we must impose a distribution on Θ.

The simplest case arises when $\mathbb{P}$ is a uniform measure - $\mathbb{P}(\omega) = 1/|\Omega|$; our goal is then to sample $\omega \in A$ uniformly. Y_1^{-1} partitions Θ into cells, each of which is a preimage of an element of A. Since $\mathbb{P}$ is uniform, Y_1^{-1} is *measure-preserving* if each cell has equal measure. In this case, sampling uniformly from Θ induces uniform samples over A. If however Y_1^{-1} is not measure preserving, the cells will differ in measure, and uniform sampling over Θ will produce biased samples in A.

Example 1. Let $\Omega = \{1, 2, 3\}, Y(\omega) = X(\omega) < 3$. Then $A = \{1, 2\}$. Let $\Theta = \{1, 2, 3, 4\}$, and $Y_1^{-1}(\theta) = 1$ if $\theta \leq 2$ else 2. Then, sample θ with probability 0.25. Alternatively, if $Y_1^{-1}(\theta) = 1$ if $\theta \leq 3$ else 2, then sample θ as 4 probability 0.5 and $1, 2, 3$ each with probability $0.5/3$

Generative Variational Family

To counteract bias when Y_1^{-1} is non-measure-preserving, we posit a second sample space Φ equipped with a uniform measure $\mathbb{P}_\Phi$, and a parameterized random variable $Z : \Phi \times W \to \Theta$. These induce a variational family [2] $Q : \Phi \times W \to \Omega$, where $Q(\phi; w) = Y_1^{-1}(Z(\phi; w))$.

Let $[\theta] = \{\theta' \mid Q(\theta'; w) = Q(\theta; w)\}$ denote the equivalence class of θ'. Then $\omega = Q(\phi; w)$ - where $\phi \sim \mathbb{P}_\Phi$ - will be uniformly distributed if:

$$\mathbb{P}_\Phi([\phi]) = 1/|A| \tag{5}$$

This condition is intractable to compute for most interesting models. Instead, we explore variational objective functionals over Q whose optima coincide with satisfying 5

Uniformity Maximization

Since we expect samples in A to be uniform, we propose to use measures of uniformity on samples in A generated by Q as its variational objective. Let A^* be a multiset $\{\omega \mid \omega = Q(\phi; w), \phi \in \Phi\}$, and $D(A^*)$ denote a measure of uniformity. A^* is *uniformly distributed* if the number of duplicate instances of each element is constant. Q is measure-preserving if and only if A^* is uniformly distributed. If A^* is uniformly distributed when its D is minimal, then Q is measure preserving when the A^* it generates has minimal D.

Many uniformity measures have been proposed [8, 3, 1], With an uniformity measure which satisfies the above conditions, the general approach for inverse simulation is as follows:

Algorithm 3. Given condition $Y : \Omega \to \{0, 1\}$, construct (i) an approximate parametric inverse $\widetilde{Y}^{-1}$, (ii) an auxiliary sample space Φ, (iii) a parameterized random variable (e.g., neural network) Z, and (iv) a variational family $\widehat{Q} : \Phi \times W \to \Omega$, where $\widetilde{Q}(\phi; w) = \widetilde{Y}_1^{-1}(Z(\phi; w))$. Take n samples $\phi_i \sim \mathbb{P}_\Phi$, and compute $(\tilde{\omega}_i, e_i) = \widetilde{Q}(\phi_i; w)$. Find w such that

$$\arg\min_w D((\tilde{\omega}_i, ..., \tilde{\omega}_n) + \sum_i e_i \tag{6}$$

Minimizing terms e and $D(\tilde{\omega})$ optimize for correctness in inversion and distribution respectively.

6 Discussion

We presented a framework for parametric inversion of non-injective functions, and outlined a variational approach to conditional sampling.

References

[1] Arijit Bishnu, Sameer Desai, Arijit Ghosh, Mayank Goswami, and Subhabrata Paul. Uniformity of point samples in metric spaces using gap ratio. 2014.

[2] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. pages 1–33, 2016.

[3] Zhonggui Chen, Zhan Yuan, Yi King Choi, Ligang Liu, and Wenping Wang. Variational blue noise sampling. *IEEE Transactions on Visualization and Computer Graphics*, 18(10):1784–1796, 2012.

[4] Ij Goodfellow, J Pouget-Abadie, and Mehdi Mirza. Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, pages 1–9, 6 2014.

[5] Noah D Goodman, Vikash K Mansinghka, Daniel M Roy, Keith Bonawitz, and Joshua B Tenenbaum. Church: a language for generative models. In *Proc. of Uncertainty in Artificial Intelligence*, page 220–229, 2008.

[6] Andrew D Gordon, Thomas A Henzinger, Aditya V Nori, and Sriram K Rajamani. Probabilistic Programming. In *International Conference on Software Engineering (ICSE, FOSE track)*, pages 1–13, 2014.

[7] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. (Ml):1–14, 2013.

[8] Stuart P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.