Adaptive Moment Estimator (Adam) Optimizer in ITKv3

Release 1.0

Bhavya Ajani[1] and Aditya Bharadwaj[2]

April 6, 2019

[1]Samsung Research Institute – Bangalore, India

Abstract

This document describes an ITK class implementing an Adaptive Moment Estimator (Adam) optimizer algorithm within the Insight Toolkit ITK www.itk.org. Adam is an adaptive gradient descent optimizer, which independently adaptively estimates the gradient descent step for each parameter, at each iteration, based on stored past gradients. The optimizer stores exponentially decaying averages of past gradients to estimate first moment (the mean) and the second moment (the variance) of the gradients to formulate update rule for present iteration. The Adam optimizer compares favorably to other adaptive learning-method algorithms, converges faster, and is robust to saddle point. This paper is accompanied with the source code, input data, parameters and output data that the authors used for validating the algorithm described in this paper.

Contents

1 Adaptive Moment Estimator (Adam) Optimizer

Image registration is a task of estimating spatial transform to align one or more images together. ITK provides a modular and pluggable registration framework to support various registration tasks [1]. Figure 1; depicts the schematic of overall registration framework implementation in ITK v3. The framework supports large number of built-in classes for most widely used components such as optimizers, metrices, transforms and interpolators. Further, these classes can be extended to support a new registration component.

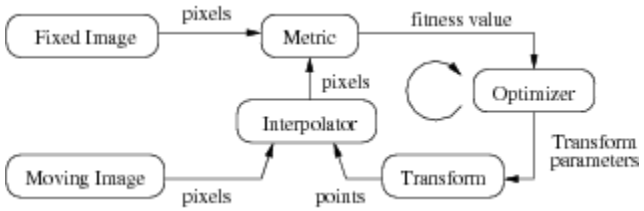

Figure 1 The basic components of the registration framework are two input images, a transform, a metric, an interpolator and an optimizer.

Optimizer is a key component of registration framework, which encapsulates the rules on updating the parameters of a transform based on current metric value and gradients. Gradient descent is one of the most popular and most commonly used optimizer to solve registration problem. ITK framework provides various gradient descent optimizers such as itk::GradientDescentOptimizer, itk::RegularStepGradientDescentOptimizer, itk::VersorRigid3DTransform Optimizer etc.

However, recently a family of gradient descent optimizers based on adaptive methods, which uses an adaptive rule for parameter update, have become popular due to faster convergence and robustness to gradient noise. One such optimizer is an Adaptive Moment Estimator or Adam optimizer. Adam optimizer is the default choice in many registration problem, especially higher dimension parametric space, due to less oscillations and smoother convergence. In this paper, we introduce newer Adam optimizer class in ITK v3 framework.

2 Implementation within ITK v3 Registration framework

Presently we introduce an Adam optimizer as a new class, itk::AdamOptimizer, within ITK v3 framework. In future, we also plan to introduce the same within newer ITK v4 registration framework.

This class extends from itk::RegularStepGradientDescentBaseOptimizer and override core functions to implement the parameter update rule as shown in Eq. 1, which is as per algorithm described in original paper titled “A Method for Stochastic Optimization” by Kingma at. el. [2].

$$\emptyset_{t+1} = \emptyset_t - \frac{\alpha}{\sqrt{\widehat{\vartheta_t} + \epsilon}} * \widehat{\omega_t} \qquad (1)$$

Here, $\emptyset_{t+1}$ is the vector of updated parameters at iteration '$t+1$' given parameters at iteration 't'. 'α' is the learning rate (which is constant) and vectors '$\widehat{\omega_t}$' and '$\widehat{\vartheta_t}$' are bias-corrected estimates of the exponentially weighted average of past gradients 'g_t' and past squared gradients as given by equation (2) and (3) respectively.

$$\widehat{\omega_t} = \frac{\omega_t}{1-\beta_1^t},\ \omega_t = \beta_1\omega_{t-1} + (1-\beta_1)g_t \quad (2)$$

$$\widehat{\vartheta_t} = \frac{\vartheta_t}{1-\beta_2^t},\ \vartheta_t = \beta_2\vartheta_{t-1} + (1-\beta_2)g_t^2 \quad (3)$$

The parameters 'β_1' and 'β_2' of the class controls amount of exponentially weighted averaging of the past gradient and squared gradients respectively.

This class overrides following virtual functions in base class to implement logic:

a) `StartOptimization`

b) `AdvanceOneStep`

c) `StepAlongGradient`

This class introduce following tunable parameters:

a) B1 (β_1) – Parameter for exponential averaging of gradient first moment, default 0.9

b) B2 (β_2)– Parameter for exponential averaging of gradient second moment, default 0.999

c) LearningRate (α) – Learning rate, default 0.1

This class does not support following parameters from the base class itk:: RegularStepGradientDescentBaseOptimizer

a) `MaximumStepLength`
b) `MinimumStepLength`
c) `RelaxationFactor`
d) `CurrentStepLength`

3 Usage

To use this class we need to include appropriate header file “itkAdamOptimizer.h” in the code.

```
#include "itkAdamOptimizer.h"
…
int main(int argc, char **argv)
{
    const unsigned int SpaceDimension = 2;
    typedef itk::Image< float, SpaceDimension > ImageType;
    typedef itk::ImageFileReader< ImageType >     ReaderType;
    //Import fixed and moving image for registration
    ….
    //The following typedef for the filter is required.
    typedef itk::AdamOptimizer       OptimizerType;
    OptimizerType::Pointer      optimizer = OptimizerType::New();
    //Generic code to create registration pipeline using ITK v3
    ….
    //Set specific parameters for this optimizer
    optimizer->SetB1(0.9); //Set parameter β1
    optimizer->SetB2(0.99); //Set parameter β2
    optimizer->SetLearningRate(1.0); //Set parameter α
    optimizer->SetNumberOfIterations(20);
    optimizer->Update();
    //Get the final parameters of the transform
    OptimizerType::ParametersType finalParameters = registration->GetLastTransformParameters();
    }
```

4 Results

The class itk::AdamOptimizer can be tested by running AdamOptimizerV3.cxx with given two input images, fixed (FixedImage.mhd) and a moving (MovingImage.mhd) image respectively. AdamOptimizerV3.cxx defines a classic ITK v3 registration pipeline to solve a 2D B-Spline deformable registration using itk::AdamOptimizer as optimizer, itk::MeanSquaresImageToImageMetric as cost function, itk:: LinearInterpolateImageFunction as interpolar, itk:: ImageRegistrationMethod as registration method and an itk:: BSplineTransform as B-Spline transform of order 3 and grid size of 7x7. The resampled registered image is saved as Registered.mhd.

Usage: ./ AdamOptimizerV3.exe FixedImage.mhd MovingImage.mhd

The fixed and moving image is of size 640 x 740 with a non-identity direction cosine. The input images are signed distance map of a segmentation of an anatomy on two different imaging modality. The registration converges within 20 iterations with final metric value of < 2.0 (with unit of mm2 given MeanSquareImageToImage metric on a signed distance images). The results are as shown below:

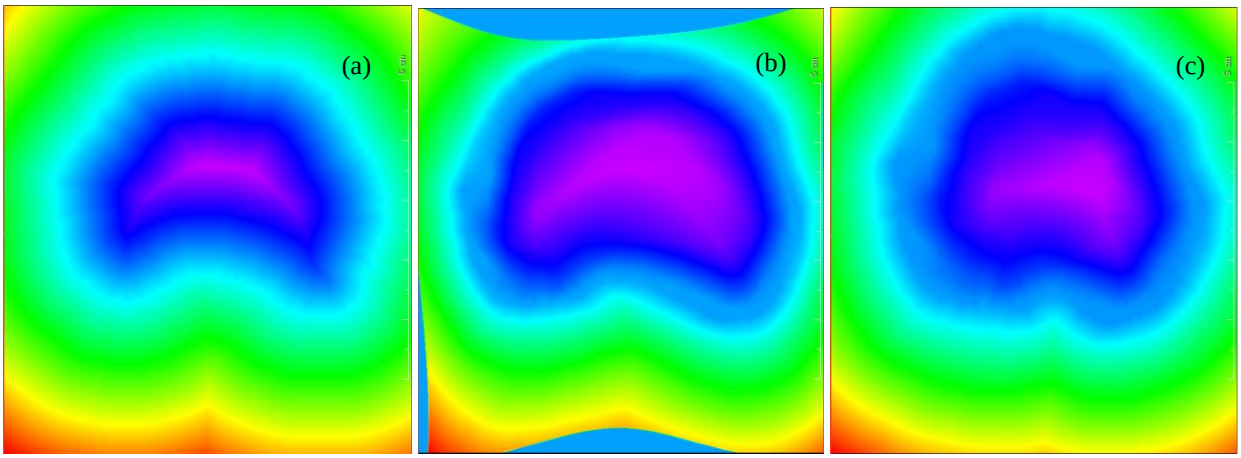

Figure 2 The results of registration using adam optimizer. (a) Shows the fixed image, (b) shows the registered image (using 2D B-Spline deformable) and (c) shows the moving image. Fixed and moving image are signed distance map of segmentation of similar anatomy on two different imaging modality.

Reference

[1] https://itk.org/ITKSoftwareGuide/html/Book2/ITKSoftwareGuide-Book2ch3.html

[2] Kingma, D. P., & Ba, J. L. (2015). Adam: a Method for Stochastic Optimization. International Conference on Learning Representations.