An Efficient Preconditioner for Stochastic Gradient Descent Optimization of Image Registration

Yuchuan Qiao, Boudewijn P. F. Lelieveldt, and Marius Staring

***Abstract*— Stochastic gradient descent (SGD) is commonly used to solve (parametric) image registration problems. In the case of badly scaled problems, SGD, however, only exhibits sublinear convergence properties. In this paper, we propose an efficient preconditioner estimation method to improve the convergence rate of SGD. Based on the observed distribution of voxel displacements in the registration, we estimate the diagonal entries of a preconditioning matrix, thus rescaling the optimization cost function. The preconditioner is efficient to compute and employ and can be used for mono-modal as well as multi-modal cost functions, in combination with different transformation models, such as the rigid, the affine, and the B-spline model. Experiments on different clinical datasets show that the proposed method, indeed, improves the convergence rate compared with SGD with speedups around 2∼5 in all tested settings while retaining the same level of registration accuracy.**

***Index Terms*— Optimization, preconditioning, stochastic gradient descent, image registration.**

I. Introduction

IMAGE registration is widely used in medical image analysis and has ample applications, e.g. in radiation therapy and segmentation [1]. This procedure can be used to align images from different modalities or different time points following a continuous deformation strategy. The strategy can be formulated as a (parametric) optimization problem to minimize the dissimilarity between a d-dimensional fixed image I_F and moving image I_M:

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}(I_F, I_M \circ \boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\mu})), \tag{1}$$

Manuscript received January 3, 2019; accepted February 3, 2019. Date of publication February 11, 2019; date of current version October 1, 2019. This work was supported in part by ZonMw, the Netherlands Organization for Health Research and Development, under Grant 104003012, and in part by the China Scholarship Council under Grant 201206130066. *(Corresponding author: Marius Staring.)*

Y. Qiao is with the Division of Image Processing, Department of Radiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands.

B. P. F. Lelieveldt and M. Staring are with the Division of Image Processing, Department of Radiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, and also with the Intelligent Systems Department, Faculty of EEMCS, Delft University of Technology, 2600 GA Delft, The Netherlands (e-mail: m.staring@lumc.nl).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2019.2897943

in which $\boldsymbol{x}$ is an image coordinate and $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\mu})$ is a coordinate transformation parameterized by $\boldsymbol{\mu}$. For example, $\boldsymbol{\mu}$ consists of rotations and translations for a rigid transformation model, and control point displacements for a nonrigid transformation modeled by B-splines. For several clinical applications, for example online adaptive radiation therapy [2], image registration runtime is crucially important. In particular, online adaptive intensity-modulated proton therapy (IMPT) [3] is very sensitive to treatment-related uncertainties, such as patient set-up, inter-fraction and intra-fraction variations in the shape and position of the target volume and organs at risk. These uncertainties should be tackled at each treatment fraction by re-optimizing the treatment plan based on a new CT scan-of-the-day. Re-contouring of the daily CT scan can be done by propagating the contour from the planning CT scan according to the spatial correspondence obtained by image registration. The registration should be performed within the time span that new organ motion occurs (less than 30 seconds for the prostate [4]), especially when a small margin is applied. A computationally efficient optimization strategy for image registration, that yields high accuracies at the same time, is therefore required.

An iterative optimization scheme is typically used:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \gamma_k \boldsymbol{D}_k, \tag{2}$$

where k is the iteration number, γ_k is the step size at iteration k, and $\boldsymbol{D}_k$ is a search direction in the parameter space. Commonly used methods to determine the search direction $\boldsymbol{D}_k$ are of first order (gradient descent) or second order (Newton or quasi-Newton) descent type. Gradient descent, however, only achieves a sublinear convergence rate for nonconvex problems or a linear convergence rate for convex problems [5], [6]. Especially for badly scaled problems, these methods converge slowly. A common example of a badly scaled problem is a rigid registration where the translational parameters can have a magnitude in the order of 1-50 mm, while the rotational parameters typically have a magnitude $\ll 1$. Second order derivative methods such as the quasi-Newton method converge faster, however, the computation of the Hessian matrix update is very time consuming, especially when the number of image voxels and transformation parameters are large [7]. For registration problems with a