A Simple Regularizer for B-spline Nonrigid Image Registration That Encourages Local Invertibility

Se Young Chun, Student Member, IEEE, and Jeffrey A. Fessler, Fellow, IEEE

Abstract—Nonrigid image registration is an important task for many medical imaging applications. In particular, for radiation oncology it is desirable to track respiratory motion for thoracic cancer treatment. B-splines are convenient for modeling nonrigid deformations, but ensuring invertibility can be a challenge. This paper describes sufficient conditions for local invertibility of deformations based on B-spline bases. These sufficient conditions can be used with constrained optimization to enforce local invertibility. We also incorporate these conditions into nonrigid image registration methods based on a simple penalty approach that encourages diffeomorphic deformations. Traditional Jacobian penalty methods penalize negative Jacobian determinant values only at grid points. In contrast, our new method enforces a sufficient condition for invertibility directly on the deformation coefficients to encourage invertibility globally over a 3-D continuous domain. The proposed penalty approach requires substantially less compute time than Jacobian penalties per iteration.

Index Terms—B-splines, local invertibility sufficient conditions, nonrigid image registration, penalty method, thorax CT images.

I. INTRODUCTION

MAGE registration is a core tool in many medical imaging applications, including fusion of structural and functional images. Several image-reconstruction schemes for MR, PET, and CT incorporate motion correction or joint estimation of motion into the reconstruction process to improve image quality [1]–[6]. Radiation treatments may be able to target cancer cells more accurately through motion correction [7]. Rigid or affine transformations can provide fast image registration. However, most of the human body does not conform to rigid or affine approximations [8]. Lamare et al. [3] used affine image registration for respiratory motion correction, but reported that it was sufficient only for a single organ and associated lesions. Effective motion correction usually requires nonrigid image registration, which enables more flexible matching of local details between two images than rigid registration.

There are many methods for nonrigid image registration [8], [9]. B-spline bases are used frequently for nonrigid image registration [10] because locally supported basis function expansions are convenient computationally and B-splines have the properties of smoothness, compact support, fast interpolation schemes, and hierarchical structure for multi-resolution [9], [11], [12]. However, deformations with high degrees of freedom can lead to unrealistic transformation results such as folding in the absence of appropriate constraints [8].

There have been some efforts to regularize nonrigid image registration based on B-splines by making certain reasonable assumptions. Rueckert et al. [13] penalized the bending energy of the deformation directly, assuming that the local deformation of tissues should be smooth. Sorzano et al. [14] proposed a regularizer based on the gradients of the divergence and the curl of the displacement field. Rohlfing et al. [15] used an incompressibility constraint: the Jacobian determinant of a transformation should be unity, assuming that local deformations are volume preserving. They applied this method after making an initial affine transformation. See [9] for other methods for constraining the transformation.

Another reasonable constraint is to impose local invertibility of the nonrigid transformation to ensure that image registration is topology-preserving or diffeomorphic.

One way to ensure local invertibility is to require the Jacobian determinant of the transformation to be positive everywhere, either as a hard constraint or by a penalty method [16]. However, most such approaches constrain the Jacobian determinant of a transformation only at each discrete voxel grid point, so local invertibility is not strictly guaranteed on the whole continuous domain. Recently, Sdika [17] described a condition involving the gradient of the Jacobian determinant that encourages the local invertibility to be achieved everywhere even though that condition is invoked only at each discrete grid point. However, compared to unregularized image registration, calculating the Jacobian determinant or its gradient significantly increases computation time due to additional B-spline interpolations of the partial derivatives of a deformation.

Ensuring invertibility is somewhat easier when using first-order B-spline bases for deformations. Musse et al. [18] derived elegant linear constraints that provide necessary and sufficient conditions to ensure that the Jacobian determinant values of such transformations are positive everywhere. However, that 2-D approach was restricted to first-order B-spline deformations. Karacali et al. [19] proposed a method to regularize 2-D and 3-D deformations to ensure that first-order B-splines are topology-preserving. Noblet et al. [20] generalized [18] for 3-D B-spline deformations and illustrated their method with first-order B-splines, but enforcing the constraints requires much higher computation than regularization based on bending energy.

Lastly, one can ensure local invertibility by imposing sufficient conditions that are simpler than the necessary conditions.
Choi et al. [21] suggested box constraints for cubic B-spline deformation coefficients that ensure invertibility, but those sufficient conditions preclude large deformations. Rueckert et al. [22] concatenated many transformations based on those box constraints to achieve large deformations. Rohde et al. [23] suggested a sufficient condition for local invertibility, derived using Neuman series for a transformation model that uses a sum of deformations. Motivated by [23], Kim et al. [24]–[26] suggested similar sufficient conditions for 3-D transformations based on cubic B-splines and implemented a constrained minimization algorithm using Dykstra’s cyclic projection method. We recently extended Kim’s sufficient conditions for local invertibility of deformations so that we can use nth-order B-spline bases; and we can also assign the upper bound on the Jacobian determinant value independently from the lower bound choice. We implemented it with a simple and fast quadratic-like penalty function [27].

This paper elaborates on the method in [27] and compares it empirically with methods based on other sufficient conditions as well as with the traditional Jacobian penalty method that uses a discrete grid [16], [25]. This paper is organized as follows. Section II reviews some related work. Section III proposes a new simple sufficient condition for the local invertibility of transformations based on B-splines and compares it with the box constraint [21], [22] empirically. Section IV proposes a new simple regularizer based on the local invertibility sufficient condition and presents 2-D and 3-D results.

II. BACKGROUND

A. Mathematical Model for Nonrigid Transformation

A 3-D nonrigid transformation \( T : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) can be written

\[
\mathcal{T}(\mathbf{z}) = \mathbf{z} + \mathbf{d}(\mathbf{z})
\]

(1)

where \( \mathbf{z} = (x, y, z) \) and \( \mathbf{d}(\mathbf{z}) \) is the deformation. We model the 3-D deformation \( \mathbf{d} = (d_x, d_y, d_z) \) using a tensor product of nth-order B-splines as follows:

\[
d^\beta(\mathbf{r}; \mathbf{q}) = \sum_{i,j,k} d_{i,j,k}^\beta \left( \frac{x}{m_x} - i \right) \beta \left( \frac{y}{m_y} - j \right) \beta \left( \frac{z}{m_z} - k \right)
\]

(2)

where \( q \in \{x, y, z\}, m_q \) is knot spacing in q direction and \( \beta \) is a nth-order B-spline basis.

The goal in image registration is to estimate the deformation coefficients \( \mathbf{d} = \{d_{i,j,k}^\beta\} \) by maximizing a similarity metric \( \Psi \):

\[
\hat{\mathbf{d}} = \arg \max_{\mathbf{d}} \Psi(\mathbf{d}; f(\mathcal{T}(\mathbf{z}; \mathbf{d})))
\]

(3)

where \( f(\mathbf{z}) \) and \( f(\mathcal{T}(\mathbf{z}; \mathbf{d})) \) denote two 3-D images.

To help stabilize the estimation, and to have physically plausible deformations, often we would like to ensure that the estimated coefficients \( \hat{\mathbf{d}} \) correspond to a diffeomorphic transformation \( T \). The methods in this paper are applicable to any similarity metric; for a survey of such metrics, see [28]. Section IV focuses on the \( l_2 \) similarity metric for registering thorax CT images at different inhalations for the purpose of radiation therapy planning and monitoring.

B. Invertibility and Diffeomorphic Transformations

Invertibility of a nonrigid transformation \( T \) is a necessary condition for it to be diffeomorphic. \( T \) is diffeomorphic if both \( T \) and \( T^{-1} \) are continuously differentiable. If we use a B-spline basis with \( n \geq 2 \) in (1), then \( T \) is continuously differentiable. (Musse et al. [18] addressed the case where \( n = 1 \).) By the implicit function theorem, if the Jacobian matrix of \( T \), denoted \( \nabla T \), is invertible everywhere, then near every point there exists a unique continuously differentiable local inverse. The determinant of the Jacobian for \( T \), denoted \( |\nabla T| \), must be non-zero for diffeomorphic nonrigid image registration. Also for \( T \) to be orientation preserving, we want \( |\nabla T| > 0 \).

Unfortunately, the condition \( |\nabla T| > 0 \) everywhere does not by itself ensure that \( T \) is globally one-to-one. One way to ensure that \( T \) is invertible globally is to ensure that transformation maps the boundary of the domain onto itself [18], [20]. However, we do not enforce such boundary conditions in this paper because the field of view for thorax inhale and exhale CT images does not contain the whole body and there is usually missing anatomy in the superior-inferior directions.

C. Related Work

The goal of diffeomorphic nonrigid image registration with the parametric representation of deformation (2) is to maximize the similarity metric (3) subject to the constraint

\[
\zeta \in C_0 \triangleq \{ \zeta : |\nabla \mathcal{T}(\mathbf{z}; \zeta)| > 0, \forall \mathbf{z} \in \mathbb{R}^3 \}.
\]

(4)

In general, this is an impractical constraint except when using linear deformation models [18]–[20] because \( \zeta \in \mathbb{R}^3 \) so there are uncountably many conditions. One way to simplify (4) is to replace the “\( \forall \mathbf{z} \)” requirement with a set of voxel grid points [16], [17], as follows:

\[
C_1 \triangleq \{ \zeta : |\nabla \mathcal{T}(\mathbf{z}; \zeta)| > 0, \forall \mathbf{z} \in \text{grid points} \}.
\]

(5)

However, because \( C_0 \subset C_1 \), this does not guarantee local invertibility between grid points. Nevertheless the smoothness of B-spline bases helps regularize \( C_1 \) so using the constraint \( C_1 \) often provides fairly good results [16]. However, computing \( |\nabla \mathcal{T}(\mathbf{z}; \zeta)| \) at all the grid points is computationally expensive.

Simplifying the condition \( |\nabla \mathcal{T}(\mathbf{z}; \zeta)| > 0 \) over \( \mathbb{R}^3 \) always involves smaller sets than \( C_0 \). Choi et al. [21] found box constraints for cubic B-spline deformation coefficients that ensure invertibility:

\[
C_2 \triangleq \{ \mathbf{d} : |d_{i,j,k}^\beta| < m_q/K, \forall i, j, k \}
\]

(6)

where \( K \approx 2.05 \) in 2-D and \( K \approx 2.48 \) in 3-D. The set \( C_2 \) provides a sufficient condition for local invertibility because \( C_2 \subset C_0 \). However, \( C_2 \) is a very restrictive constraint set that allows only very small deformations. To achieve large deformations, Rueckert et al. [22] composed several transformations that each satisfied this condition.

Kim et al. [24]–[26] suggested a sufficient condition for ensuring invertibility of cubic B-spline deformations that allows
a larger family of deformations. Instead of restricting the absolute values of the coefficients as in (6), this condition limits the differences of adjacent B-spline coefficients

\[
C_3 \triangleq \bigcap_{q \in \mathbb{Q} \setminus \{x, y, z\}} \left\{ \varepsilon : \left| c_{i,j,k}^q - c_{i,j,k}^q \right| < m_{kq}^2, \right\}
\]

where \(k_x + k_y + k_z < 1\). Although \(C_3 \subset C_0\), this sufficient condition only allows large deformations with fairly small Jacobian determinant values. In particular, one can show that

\[
1 - (k_x + k_y + k_z) \leq |\nabla \mathcal{L}(x, y, z)| \leq (1 + k_x)(1 + k_y)(1 + k_z) + (1 + k_x)k_yk_z + k_x(1 + k_y)k_z + k_xk_y(1 + k_z) \quad \forall q \in C_3 [24]-[26].
\]

This means that \(C_3\) does not allow acute volume changes locally. This is because the upper bound on the Jacobian determinant is determined by the lower bound design. For example, if we choose \(k_q = 1/3\) so that the lower bound for the Jacobian determinant \(|\mathbf{J}|\) is 0, then the upper bound for the Jacobian determinant value would be automatically determined to \(76/27 \approx 2.8148\) which is fairly small [25]. The next section provides new broader sets of sufficient conditions.

### III. LOCAL INVERTIBILITY CONDITION

#### A. Lemmas

We first extend Kim’s sufficient conditions for local invertibility to overcome two limitations [27]. Firstly, a 7th-order B-spline basis \((n \geq 1)\) can be used instead of cubic B-spline basis for deformation modeling. Secondly, the upper bound of Jacobian determinant can be designed independently from the lower bound of Jacobian determinant.

**Lemma 1:** For concise notation, denote the Jacobian \(\mathbf{J} = \nabla \mathcal{T}\) of a 3-D transformation as

\[
\mathbf{J} = \mathbf{I} + \begin{pmatrix}
  x_1 & x_2 & x_3 \\
  x_4 & x_5 & x_6 \\
  x_7 & x_8 & x_9
\end{pmatrix}.
\]

Then the corresponding determinant is given by

\[
|\mathbf{J}| = (1 + x_1)(1 + x_2)(1 + x_3) + x_2x_3x_7 + x_3x_4x_8 - (1 + x_1)x_6x_8 - (1 + x_5)x_3x_7 - (1 + x_9)x_2x_4.
\]

Suppose that the elements of the 3-D Jacobian determinant satisfy \(x_1 \in I_i, i = 1, \ldots, 9\) where \(I_i \subset \mathbb{R}\) are compact intervals. Then \(|\mathbf{J}|\) achieves its global maximum and minimum values over \(I = I_1 \times \cdots \times I_9\) and those maximum and minimum values are achieved for a point \(x_i^*\) for which \(x_i^* \in \{\max I_i, \min I_i\}\) for \(\forall i = 1, \ldots, 9\).

The Appendices have the proofs of these Lemmas. This Lemma implies that we can determine the global minimum and maximum of \(|\mathbf{J}|\) over the compact set \(\mathcal{I}\) “simply” by calculating the \(2^9\) possible values of \(|\mathbf{J}|\) at the vertices of \(\mathcal{I}\). (It is trivial to apply this Lemma to 2-D cases.)

Kim *et al.* provided a specific formula for the “possible” maximum and minimum of \(|\mathbf{J}|\) for given ranges of each \(x_i\) value using Karush–Kuhn–Tucker conditions [25]. We suggest next a generalization using Lemma 1.

**Lemma 2:** Suppose that \(|x_i| \leq k_iq < 1/2\) where \(q_i = \text{for } i = 2, 3, q_i = y \text{ for } i = 4, 6 \text{ and } q_i = z \text{ for } i = 7, 8\). Also suppose that \(k_i \leq |x_i| \leq k_{pi} \text{ where } p_i = \text{for } i = 1, p_i = y \text{ for } i = 5 \text{ and } p_i = z \text{ for } i = 9\). Then \(\min |\mathbf{J}| = 1 - (k_x + k_y + k_z)\) and \(\max |\mathbf{J}| = (1 + k_x)(1 + k_y)(1 + k_z) + (1 + k_x)k_yk_z + k_x(1 + k_y)k_z + k_xk_y(1 + k_z)\). In other words

\[
1 - (k_x + k_y + k_z) \leq |\mathbf{J}| \leq (1 + k_x)(1 + k_y)(1 + k_z) • (1 + k_x) + (1 + k_y)k_yk_z + k_x(1 + k_y)k_z + k_xk_y(1 + k_z).
\]

Kim’s proposition was restricted to the case where \(k_x = k_y = k_z\). To ensure local invertibility, \(k_x + k_y + k_z\) should be less than 1, where each \(k_q\) is positive, so that the lower bound in (9) is positive.

Kim *et al.* showed a second proposition about the relationship between the first partial derivative of deformation and adjacent deformation coefficients for the cubic B-spline basis case [25]. We show next that this relation is also valid for general \(n\)-th-order B-spline bases \((n \geq 1)\). We also generalize the bounds used by Kim *et al.* with Lemma 2 [27].

**Lemma 3:** If \(b_i \leq c_{i+1,j,k}^p - c_{i,j,k}^p \leq b_M\) for \(\forall i, j, k\), then \(b_i / m_x \leq \partial \mathcal{L}(q) / \partial x \leq b_i / m_x\) for \(\forall q \in \mathcal{I}\) where \(q \in \{x, y, z\}\).

Similarly, if \(b_i \leq c_{i,j+1,k}^p - c_{i,j,k}^p \leq b_M\) for \(\forall i, j, k\), then \(b_i / m_y \leq \partial \mathcal{L}(q) / \partial y \leq b_i / m_y\) and if \(b_i \leq c_{i,j,k+1}^p - c_{i,j,k}^p \leq b_M\) for \(\forall i, j, k\), then \(b_i / m_z \leq \partial \mathcal{L}(q) / \partial z \leq b_i / m_z\) for \(\forall q\) respectively.

This Lemma limits the range of values of the first derivative of \(\partial \mathcal{L}(q) / \partial \mathbf{R}^3\) by restricting the differences of adjacent deformation coefficients. Combined, Lemmas 2 and 3 show that one can obtain a transformation \(\mathcal{T}\) that is everywhere locally invertible by maximizing a similarity metric subject to constraints on the differences between adjacent deformation coefficients, as summarized in the following Theorem.

**Theorem 1:** Suppose \(0 \leq k_q < 1/2\) for \(q \in \{x, y, z\}\). Define

\[
C_4 \triangleq \left\{ \varepsilon : -m_x k_x e_i^p - e_{i+1,j,k}^p - e_{i,j,k}^p \leq m_x K_x, -m_y k_y e_i^p - e_{i,j+1,k}^p - e_{i,j,k}^p \leq m_y K_y, -m_z k_z e_i^p - e_{i,j,k+1}^p - e_{i,j,k}^p \leq m_z K_z, \right\}
\]

In (2), if \(c \in C_4 \) then \(|\mathbf{J}|\) satisfies the bounds in (9) \(\forall q \in \mathcal{I}\). Moreover, if \(k_x + k_y + k_z < 1\), then the transformation (2) is locally invertible everywhere.

This theorem applies to deformations based on any \(n\)-th-order B-spline basis. We set the lower and upper bounds for \(|\mathbf{J}|\) by setting appropriate \(k_q\) and \(K_q\) values for \(q \in \{x, y, z\}\).
The constrained transformation maps $[0 0.6]$ to $[0.3 0.6]$ instead of $[0.3 0.6]$. This deformation belongs to $C_4$ because $-1 < \partial d(x)/\partial x < \infty$. However, if we impose the sufficient condition $-0.33 < \partial d(x)/\partial x$, then Fig. 1 shows that acute volume shrinkage is precluded because the minimum derivative of the transformation is 0.67. The constrained transformation maps $[0 0.6]$ to $[0.3 0.7]$ instead of $[0.3 0.6]$. More generally, when we choose $k_x, k_y$, and $k_z$ subject to $k_x + k_y + k_z < 1$ to ensure invertibility, $C_4$ imposes restrictions for acute volume changes in each direction.

The 2-D case illustrates the solution space of $C_4$ in terms of Lemma 2. Lemma 2 is trivial for a 2-D Jacobian determinant $|J| = (1 + a)(1 + d) - bc$ where $J = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. A deformation having a positive Jacobian determinant must satisfy $(1 + a)(1 + d) > bc$. We can introduce a free parameter $k$ such that $|J|$ is always positive if $(1 + a)(1 + d) > k$ and $bc < k$ for any $k$. Fig. 2 visualizes the solution space for 2-D invertible deformations in terms of $a, b, c, d$, and $k$. For fixed $k$, any values of $(a, d)$ that lie above the upper line or below the lower line yield a positive Jacobian determinant if $(b, c)$ lies between these lines. Lines vary as $k$ varies. To allow acute volume shrinkage, we need $k$ to be close to 0 as observed in Fig. 1. However, smaller $k$ values imply more restrictive sets for $(b, c)$.

Lemma 2 corresponds to fixing $k = k_xk_yk_z$ such that $k_x + k_y < 1$ and $k_x \geq 0, k_y \geq 0$. This yields the rectangular areas for $a, b, c, d$ shown in Fig. 3 (for $k_x = k_y = 1/2$ and $k = 1/4$). Thus, Theorem 1 not only uses a fixed value for $k$, but also imposes restrictive box constraints on the deformation derivatives. However, it still has a larger solution space than traditional box constraints on the B-spline coefficients such as [21]. Because $k$ is fairly small, relaxing this sufficient condition may allow larger volume shrinkage [27].

C. Concatenating Transformations

Since $C_4$ is a restrictive sufficient condition, it may not contain all real deformations of interest. To allow larger deformations, we can concatenate multiple elemental transformations that belong to $C_4$, i.e., let $T(x) = T_N(\cdots (T_2(T_1(x)))))$ where each $T_k$ satisfies $C_4$. Since each $T_k$ is diffeomorphic, $T$ is also diffeomorphic.

Rueckert et al. [22] used a box constraint $C_2$ [21] to guarantee that each elemental transformation is diffeomorphic. We use $C_4$ for our elemental transformations. This should require fewer elemental transformations because $C_4$ allows a larger solution space than $C_2$, as illustrated in the next section.

D. 2-D Simulation: Warping a Disk to a “C” Shape

We applied several constrained nonrigid image registration methods to the challenging registration problem shown in Fig. 4. We placed deformation knot points every 4th pixel, i.e., $m_x = m_y = 4$. The data fit term used sum of squared differences. For optimization, we used augmented Lagrangian multipliers [17] with the conjugate gradient method. Line search step size was determined by one step of Newton’s method. We used fast B-spline interpolation [29]–[31] with a 4-level multiresolution scheme [32].

Fig. 5 shows the unconstrained registration and results of using $C_1, C_2$ and $C_4$. The unconstrained result in Fig. 5(a) shows some unrealistic warping such as folding. Fig. 5(b) shows the regularized deformed images with a Jacobian penalty based on $C_1$. This shows a more regular warp than Fig. 5(a). However, $C_1$ allows a larger solution space than the ideal solution space $C_0$.

Fig. 5(c) and (d) show the limitation of using a single warp based on $C_2$ and $C_4$ respectively. The sufficient conditions $C_2$ and $C_4$ do not contain the complicated diffeomorphic transformation needed to map the source image to the target image in Fig. 4. However, this warp can be achieved satisfactorily by composing just three warps that each belong to $C_4$, as shown in Fig. 6(b). In contrast, to achieve a satisfactory warp by composing transformations that lie inside the box constraint $C_2$ [21], [22] required about 30 concatenations, as shown in Fig. 6(a).
For larger and more complicated deformations, our proposed constraint $C_4$ can be used as a simple elemental transformation to provide diffeomorphic composite transformations.

IV. SIMPLE REGULARIZER BASED ON LOCAL INVERTIBILITY CONDITION

A. Proposed Simple Regularizer

If we want to strictly ensure local invertibility, then we maximize a similarity metric subject to the linear constraints $\xi \in C_4$. However, to simplify the computation, we can relax the invertibility condition by using a penalty method [16], [27]. In a penalty method, we maximize an objective function that is the similarity metric minus a penalty function that encourages the invertibility condition, but does not enforce it strictly.

We propose to construct a penalty function based on the following piecewise quadratic function

$$p(t; \zeta_1, \zeta_2) = \begin{cases} \frac{1}{2}(t - \zeta_1)^2, & t < \zeta_1 \\ 0, & \zeta_1 \leq t \leq \zeta_2 \\ \frac{1}{2}(t - \zeta_2)^2, & \zeta_2 < t \end{cases}$$

which is illustrated in Fig. 7. The argument $t$ denotes a difference between two adjacent deformation coefficients. This function does not strictly constrain such differences, but its first and second derivatives are simple and convenient for use in optimization algorithms such as conjugate gradient. The final new penalty function is

$$R(\xi) = \sum_{q \in [x,y,z]} \sum_{i,j,k} \left[p\left(\xi_{i,j,k}^q - \xi_{i,j,k}^q; \zeta_1^q, \zeta_2^q\right) + p\left(\xi_{i,j,k+1}^q - \xi_{i,j,k}^q; \zeta_1^q, \zeta_2^q\right) + p\left(\xi_{i,j+1,k}^q - \xi_{i,j,k}^q; \zeta_1^q, \zeta_2^q\right)\right]$$

(10)

where $\xi_{i,j,k}^q = -m_q k_q$ for $\forall r \in \{x,y,z\}$, $\zeta_1^q = m_q k_q$ for $r \neq q$ and $\zeta_2^q = m_q K_q$ for $r = q$. Note that choosing $\zeta_1 = \zeta_2 = 0$ would correspond to a quadratic roughness penalty over B-spline coefficients, which is akin to encouraging the volume preserving condition $[\mathbf{I}] = 1$, $\forall \mathbf{I}$.

Being based on the somewhat restrictive solution space $C_4$, the new penalty method can encourage the local invertibility on the whole continuous domain with a fast and memory-efficient implementation. This implementation is possible because $C_4$ does not require additional B-spline interpolations beyond the interpolations needed for the data fitting term. It also encourages the smoothness of deformations inherently because it constrains the differences between adjacent deformation coefficients. In contrast, using $C_0$ or $C_1$ is much more expensive for one transformation.

B. Incorporating a Priori Knowledge of Motions

For diffeomorphic transformations using Theorem 1, the usual choice would be $k_x = k_y = k_z = 1/3 - \varepsilon$ for some small $\varepsilon$. However, if we have a priori knowledge about the deformation, then we can assign each $k_q$ accordingly. For instance, for registering thorax inhale and exhale images, we can assign $k_x = k_y = 1/4 - \varepsilon$ and $k_z = 1/2 - \varepsilon$ because the deformation in the $z$ direction is larger due to diaphragm motion, whereas the deformations in the $x$ and $y$ directions are smaller. With this design, $C_4$ allows 50% local shrinkage along $z$ and 75% local shrinkage along $x$ and $y$ instead of allowing 67% shrinkage in each direction. We can use this sufficient condition for the
proposed simple regularizer to encourage local invertibility of the deformation.

C. 2-D Simulation: Expansion and Shrinkage

We applied nonrigid image registration to the 256 × 256 images in Fig. 8 using no constraint, a Jacobian penalty based on $C_1$, a quadratic roughness penalty [7], a regularizer based on Kim’s constraint $C_3$, and our proposed penalty method based on $C_4$. Fig. 8 has an expanding circle and a shrinking ellipse to illustrate the difference between $C_3$ and $C_4$. Since we have a priori knowledge about vertical motion, we investigated two sets of parameters in $C_4$: a symmetric way with $k_x = k_y = 1/2 - 0.01 \times 1/2$ as well as an asymmetric way with $k_x = 0.05 - 0.01 \times 0.35 < k_y = 0.65 - 0.65 \times 0.01$.

We placed deformation knot points every 4th pixel. The data fit term used sum of squared differences. For optimization we used the conjugate gradient method. Line search step size was determined by one step of Newton’s method. We used fast B-spline interpolation and the 4-level multiresolution scheme as in Section III-D. We ran 200 iterations for each level or ran until the $l_0$ norm of the gradient is less than the machine accuracy. We checked the local invertibility by computing Jacobian determinant values on a grid ten times finer than the image resolution.

Fig. 9 quantifies the tradeoff between image similarity and local invertibility for the five different registration methods for a range of regularization parameters. The horizontal axis is the root mean square (rms) difference between the deformed image and the target image (log scale) and the vertical axis is the number of the finer (ten times) voxel grid points having a non-positive Jacobian determinant (log scale). We took a log after adding 1 for the number of non-positive Jacobian determinant since the lowest number of it is 0.

For the unconstrained case, the rms difference was 0.0109 and the number of negative Jacobian determinants was 497644. As the regularization parameters decrease, the rms differences and the number of negative Jacobian determinants of all other methods approached closely to these values. This is expected because the unconstrained case is the same as any other penalty method with regularization parameter 0.

As we increased the regularization parameters, the number of negative Jacobian determinants “generally” decreases, eventually towards zero, although not always monotonically. The rms differences also “generally” increase as the regularization parameters increase for most methods except Jacobian penalty. This is because the Jacobian constraint $C_3$ contains the original constraint $C_3$ and it does not restrict the deformation so that it can achieve low rms difference for strong penalty parameters. For properly chosen regularization parameters, symmetric/asymmetric proposed simple penalties show fairly good performance compared to Kim’s or quadratic penalty methods based on more restrictive sufficient conditions (Kim’s: $K_x = K_y$, $K_y = K_y$ and quadratic: $K_x = K_y = 0, K_y = K_y = 0$).

Table I shows the best rms difference of image for each method with zero non-positive Jacobian determinant values over the ten-times finer grid. The proposed simple penalty with a priori motion information performed well compared to Jacobian penalty. However, as the regularization method depends on more restrictive condition, the rms difference is
Fig. 6. Proposed constraint requires much less transformations to achieve a satisfiable deformation than the box constraint. (a) 10, 20, 30 compositions of box constraint $C_2$ and a warped grid of 30 compositions. (b) 1, 2, 3 compositions of proposed constraint $C_4$ and a warped grid of three compositions.

Fig. 7. Variant of quadratic penalty function (solid) and real constraints (dashed) used with constraint set $C_4$.

Fig. 8. Images for illustrating expansion and shrinkage. (a) 2-D source image. (b) 2-D target image.

larger. It clearly shows that a quadratic penalty oversmooths the deformations.

Our proposed asymmetric penalty performed a little better than Jacobian penalty in these experiments. However, our proposed penalty may not always perform better. It depends on the convergence, regularization parameter, image structure and so on. Since the data fitting term is non-convex, local minima may affect the result, too. However, for simpler cases like in Fig. 8 our proposed regularization method may perform close to the Jacobian penalty method. In the next section, we apply both methods to the 3-D real CT images of a patient.

D. 3-D Real CT Images

We applied our proposed regularization method (10) to the problem of registering 3-D breath-hold X-ray CT images of a real oncology patient scanned at inhale and at exhale. These images are useful for radiation treatment planning. The image size was $396 \times 256 \times 128$ as shown in Fig. 10. We chose $k_x = \frac{1}{4} - 0.01 \times \frac{1}{4}$ and $k_y = \frac{1}{2} - 0.01 \times \frac{1}{2}$ because we expect the deformation in the $z$ direction to be larger than the deformations in the $x$ and $y$ directions due to diaphragm motion.

We used the same methods as in Section IV-C except for the multiresolution scheme. For the first three levels of multiresolution, the knot spacing was every eight pixels for downsampled images, and for the last level of multiresolution the knot spacing was every four pixels. We ran 120 iterations at each level to see the convergence properties. The regularization parameter that multiplies (10) was chosen experimentally to achieve the minimum value of data fitting term such that all Jacobian values on the image grid were positive.

Fig. 11 shows the difference images between the target image and the deformed images. As expected, the difference image for unconstrained registration in Fig. 11(a) has smaller values than the constrained difference images in Fig. 11(b) and (c). The rms difference for unconstrained registration was the smallest, which was 19.9 HU. The rms errors of the Jacobian penalty (25.9 HU) and the proposed penalty (29.2 HU) were somewhat higher. However, Fig. 12(a) shows that unconstrained registration yields an unrealistic warped grid. The number of negative Jacobian determinant voxels was 316914 out of 12582912 voxels. Fig. 12(c) shows a smoother warp than Fig. 12(b) because our proposed penalty method is based on $C_4$ which is a smaller set than $C_3$. Our proposed method has smoothness property implicitly because it restricts the range of the differences between adjacent B-spline coefficients.
The proposed penalty method was much faster and more memory efficient than the traditional Jacobian penalty method per iteration. If one uses the sum of squared error as the data fitting term and penalizes negative Jacobian determinant values on each image grid point in 3-D with cubic B-splines, then the interpolations needed to compute the gradients of the direct Jacobian penalty function require about 1.8 times more operations than the interpolations needed for the gradient of the data fitting term. Table II shows the computational cost for one iteration at the last (finest) level of the multiresolution procedure. Our proposed method requires only slightly more time per iteration than unconstrained registration, and much less time than using a Jacobian penalty. Furthermore, in this simulation, the traditional Jacobian penalty method required about twice as much memory as our proposed method because it must store the interpolation results for the Jacobian gradient. Fig. 13 shows the convergence of each method.

We could compose a coarse resolution warp based on (10) with one full sequence of coarse-to-fine warp based on (10) to reduce rms differences further with only slight increase in computation.

V. DISCUSSION

We proposed a new condition Theorem 1 that is sufficient to ensure the local invertibility of transformations based on B-splines. Its limitation can be overcome by using composite transformations. This proposed sufficient condition can be used with constrained optimization such as augmented Lagrangian multiplier method [17] or Dykstra’s cyclic projection method [25].

We showed that the proposed sufficient condition is more general than other simple sufficient conditions that ensure local invertibility everywhere such as box constraint [21]. When used in composite transformations, it requires many fewer transformations to achieve comparable deformations [22].

We also relaxed our local invertibility condition by a simple quadratic-like penalty. This approach achieves more flexible image matching compared to other penalty methods based on more restrictive local invertibility conditions. For practical use in a thorax image registration, we used a single transformation with a simple quadratic-like penalty that encourages \( \xi \in C_4 \). This gave a fairly good deformation with no negative Jacobian determinant values on image voxel grid points. This approach is much simpler and faster than the traditional Jacobian determinant penalty and is more memory efficient.

Some application areas require not only local invertibility, but also require computing the inverse transformation. One approach is to estimate both forward and backward image registration parameters with consistency regularizer [33]. Using both consistency regularizer and our proposed regularizer can be interesting future work.

In the 3-D thorax registration example, we observed some bone warping in each of the deformed images; a natural direction for further research would be to use a rigidity penalty term [34], [35].

APPENDIX A

PROOF OF LEMMA 1

Proof: The global maximum and minimum values exist since \( |J| \) is continuous on the compact set \( I_1 \times \cdots \times I_9 \). Suppose that \( (x_1, \ldots, x_9) \) achieves the global minimum value of \( |J| \) and
\[ \min I_k < x_k < \max I_k \] for some \( k \). Fix all \( x_i \) except \( x_k \), \( \{J\} \) is an affine function with respect to \( x_k \) so \( \{J\} \) can achieve equal or better global minimum value on either \( x_k = \min I_k \) or \( x_k = \max I_k \). The same argument can be applied for all \( x_i \) such that \( \min I_i < x_i < \max I_i \) and thus it generates a contradiction. The same argument can be applied to the global maximum case.

**APPENDIX B**

**PROOF OF LEMMA 2**

Proof: By Lemma 1, we need to evaluate \( \{J\} \) only on \( x_1 \in \{ -k_x, K_x \} \), \( x_4 \in \{ -k_y, K_y \} \), \( x_9 \in \{ -k_z, K_z \} \) and \( x_i \in \{ -k_i, k_i \} \) where \( q_i = x \) for \( i = 2, 3, q_i = y \) for \( i = 4, 6 \) and \( q_i = z \) for \( i = 7, 8 \). For fixed \( x_i \) except \( x_1 \), \( \{J(x_1)\} = (1 + x_1)\{(1 + x_2)(1 + x_3) - x_6x_8\} + c \) where \( c \) is a constant for \( x_1 \) and \( (1 + x_2)(1 + x_3) - x_6x_8 \) is always positive under given conditions. So \( x_1 = K_x \) for \( \max \{J\} \) and \( x_1 = -k_x \) for \( \min \{J\} \). Similarly we determine \( x_5 \) and \( x_9 \). For fixed \( x_i \) except \( x_2 \), \( \{J(x_2)\} = x_2\{x_3x_7 + (1 + x_3)x_4\} + c \) where \( c \) is a constant for \( x_2 \). For \( \min \{J\} \), \( x_2 = -k_x \) if \( x_4 = -k_y \) and \( x_2 = k_x \) if \( x_4 = k_y \). In other words, \( x_2x_4 = k_xk_y \) for \( \min \{J\} \). Similarly, \( x_3x_4 = -k_xk_y \) for \( \max \{J\} \). In this fashion, \( x_6x_8 \) and \( x_3x_7 \) will be determined for \( \max \{J\} \) and \( \min \{J\} \). From these results, one can induce that \( x_2x_4x_7 + x_3x_4x_8 = 0 \) for \( \max \{J\} \) and \( x_2x_4x_7 = x_3x_4x_8 = -k_xk_yk_z \) for \( \min \{J\} \).
Using the constraints \( b_m \leq c_{i+1,j,k} - c_{i,j,k} \leq b_M \) and the property \( \sum_i \beta^\alpha(x/m_x - i) = 1 \), we have the bounds

\[
\frac{\partial}{\partial x} d(x) = \sum_i c_i \frac{\partial}{\partial x} \beta^\alpha(x/m_x - i)
\]

\[
= \sum_i (c_i - c_{i-1}) \beta^\alpha-1(x/m_x - i + 1/2)/m_x.
\]

Similarly, \( \partial d(x)/\partial x \geq b_m/m_x \). The other directions \( y, z \) can be proved similarly.

**ACKNOWLEDGMENT**

The authors thank Dr. M. Kessler for the 3-D CT data and Dr. J. Balter for discussions about image registration. We also thank the reviewers for insightful questions that improved the paper.

**REFERENCES**


