Magnetic Resonance in Medicine 68:730–740 (2012)

Nonlinear GRAPPA: A Kernel Approach to Parallel MRI Reconstruction

Yuchou Chang,[1] Dong Liang,[1,2] and Leslie Ying[1]*

GRAPPA linearly combines the undersampled *k*-space signals to estimate the missing *k*-space signals where the coefficients are obtained by fitting to some auto-calibration signals (ACS) sampled with Nyquist rate based on the shift-invariant property. At high acceleration factors, GRAPPA reconstruction can suffer from a high level of noise even with a large number of auto-calibration signals. In this work, we propose a nonlinear method to improve GRAPPA. The method is based on the so-called kernel method which is widely used in machine learning. Specifically, the undersampled *k*-space signals are mapped through a nonlinear transform to a high-dimensional feature space, and then linearly combined to reconstruct the missing *k*-space data. The linear combination coefficients are also obtained through fitting to the ACS data but in the new feature space. The procedure is equivalent to adding many virtual channels in reconstruction. A polynomial kernel with explicit mapping functions is investigated in this work. Experimental results using phantom and in vivo data demonstrate that the proposed nonlinear GRAPPA method can significantly improve the reconstruction quality over GRAPPA and its state-of-the-art derivatives. Magn Reson Med 68:730–740, 2012. © 2011 Wiley Periodicals, Inc.

Key words: parallel imaging; GRAPPA; nonlinear filtering; reproducible kernel Hilbert space; kernel method; regularization

Among many partially parallel acquisition methods (e.g., 1–9), generalized auto-calibrating partially parallel acquisitions (GRAPPA) (1) has been widely used for reconstruction from reduced acquisitions with multiple receivers. When the net acceleration factor is high, GRAPPA reconstruction can suffer from aliasing artifacts and noise amplifications. Several methods have been developed in recent years to improve GRAPPA, such as localized coil calibration and variable density sampling (10), multicolumn multiline interpolation (11), regularization (12,13), iteratively reweighted least-squares (14), high-pass filtering (15), cross validation (16,17), iterative optimization (18), GRAPPA operator (19,20), virtual coil using conjugate symmetry (21), multislice weighting (22), infinite pulse response filtering (23), cross sampling (24), and filter bank methods (25,26).

[1]Department of Electrical Engineering and Computer Science, University of Wisconsin, Milwaukee, Wisconsin 53211, USA.

[2]Paul C. Lauterbur Research Center for Biomedical Imaging, Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Shenzhen, P. R. China.

*Correspondence to: Leslie Ying, Ph.D., Department of Electrical Engineering and Computer Science, University of Wisconsin, Milwaukee, Wisconsin 53211. E-mail: leiying@uwm.edu

Received 30 May 2011; revised 4 October 2011; accepted 10 October 2011.

DOI 10.1002/mrm.23279

Published online 12 December 2011 in Wiley Online Library (wileyonlinelibrary.com).

The conventional GRAPPA method (1) reconstructs the missing *k*-space data by a linear combination of the acquired data, where the coefficients for combination are estimated using some auto-calibration signal (ACS) lines usually acquired in the central *k*-space. Huang et al. (27) analyzed two kinds of errors in GRAPPA reconstruction: truncation error and inversion error. Nana et al. (16,17) extended the analysis and used more general terms: model error and noise-related error. The first kind of error mainly originates from a limited number of ACS lines and data truncation. When a limited size of *k*-space signals is observed or inappropriately chosen instead of the whole *k*-space, model errors occur in GRAPPA reconstruction. This type of error usually varies with the amount of ACS data, reduction factor, and the size of the coefficients to be estimated for reconstruction. For example, a reduction in ACS acquisition usually results in degraded image quality. Therefore, a large amount of ACS data is needed to reduce this model error but at the cost of prolonged acquisition time. The second kind of errors originates from noise in the measured data and noise-induced error in estimating the coefficients for linear combination. Regularization (12,13) has been used in solving the inverse problem for the coefficients, but significant noise reduction is usually at the cost of increased aliasing artifacts. Iterative reweighted least-squares (14) method reduces the noise-induced error to a greater extent by ignoring noise-induced "outliers" in estimating the coefficients. However, the method is computationally expensive.

In this paper, we focus on the nature of noise-induced error and develop a novel nonlinear method to reduce such kind of error. We identify the nonlinear relationship between the bias in the estimated GRAPPA coefficients and the noise in the measured ACS data due to the error-in-variable problem in the calibration step. This relationship suggests that the finite impulse response model currently used in GRAPPA reconstruction is not able to remove the nonlinear noise-induced bias even if regularization is used. We thereby propose a nonlinear approach to GRAPPA using the kernel method, named nonlinear GRAPPA. (Note this kernel is a terminology in machine learning and is different from the GRAPPA kernel for linear combination.) The method maps the undersampled data onto a high dimensional feature space through a nonlinear transform and the data in the new space are then linearly combined to estimate the missing *k*-space data. Although the relationship between the acquired and missing *k*-space data is nonlinear, the relationship can be easily and linearly found in the high dimensional feature space using the ACS data. It is worth noting that the nonlinearity of this approach is

completely different from that in the GRAPPA operator formulation in Refs. (19,20) where the former is on the *k*-space data while the latter is on the GRAPPA coefficients through successive application of linear operators. The proposed method not only has the advantage of nonlinear methods in representing generalized models that include linear ones as a special case, but also maintains the simplicity of linear methods in computation.

THEORY

Review of GRAPPA

In conventional GRAPPA, the central *k*-space of each coil is sampled at the Nyquist rate to obtain ACS data, while the outer *k*-space is undersampled by some outer reduction factors (ORF). The missing *k*-space data is estimated by a linear combination of the acquired undersampled data in the neighborhood from all coils, which can be represented mathematically as

$$S_j(k_y + r\Delta k_y, k_x) = \sum_{l=1}^{L} \sum_{t=B_1}^{B_2} \sum_{h=H_1}^{H_2} w_{j,r}(l,t,h) \times S_l(k_y + tR\Delta k_y, k_x + h\Delta k_x), j = 1, ..., L, r \neq tR, \quad [1]$$

where $S_j(k_y + r\Delta k_y, k_x)$ denotes the unacquired *k*-space signal at the target coil, $S_j(k_y + tR\Delta k_y, k_x + h\Delta k_x)$ denotes the acquired undersampled signal, and $w_{i,j}(l,t,h)$ denotes the linear combination coefficients. Here R represents the ORF, l counts all coils, t and h transverse the acquired neighboring *k*-space data in k_y and k_x directions, respectively, and the variables k_x and k_y represent the coordinates along the frequency- and phase-encoding directions, respectively.

In general, the coefficients depend on the coil sensitivities and are not known a priori. The ACS data are used to estimate these coefficients. Among all the ACS data fully acquired at the central *k*-space, each location is assumed to be the "missing" point to be used on the left-hand side of Eq. 1. The neighboring locations with a certain undersampling pattern along the phase encoding direction are assumed to be the undersampled points that are used on the right-hand side of Eq. 1. This is repeated for all ACS locations (except boundaries of the ACS region) based on the shift-invariant property to fit GRAPPA coefficients to all ACS data. This calibration process can be simplified as a matrix equation

$$\mathbf{b} = \mathbf{A}\mathbf{x}, \quad [2]$$

where $\mathbf{A}$ represents the matrix comprised of the undersampled points of the ACS, $\mathbf{b}$ denotes the vector for the "missing" points of the ACS, and $\mathbf{x}$ represents the coefficients to be fitted. The matrix $\mathbf{A}$ is of size $M \times K$ with M being the total number of ACS data (excluding the boundaries) and K being the number of points in the neighborhood from all coils that are used in reconstruction. The least-squares method is commonly used to calculate the coefficients:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2. \quad [3]$$

When the matrix $\mathbf{A}$ is ill-conditioned, the noise can be greatly amplified in the estimated coefficients. To address the ill-conditioning issue, regularization methods (12,13) have been used to solve for coefficients using a penalized least-squares method,

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \lambda R(\mathbf{x}), \quad [4]$$

where $R(\mathbf{x})$ is a regularization function (e.g., $R(\mathbf{x}) = ||\mathbf{x}||_2$ in Tikhonov) and λ is regularization parameter. Regularization can effectively suppress noise to a certain level. However, aliasing artifacts usually appear in reconstruction at the same time while large noise is suppressed.

Another source of noise-induced error in GRAPPA is "outliers." Outliers are *k*-space data with large measurement errors due to noise and low sensitivity. They lead to large deviations in the estimated coefficients from the true ones when the least squares fitting is used. Iterative reweighted least-squares method (14) has been proposed to minimize the effect of outliers in least-squares fitting. The method iteratively assigns and adjusts weights for the acquired undersampled data. "Outliers" are given less weights or removed in the final estimation, so that the fitting accuracy and reconstruction quality are improved. However, the high computational complexity of the method limits its usefulness in practice.

Errors-in-Variables Model of GRAPPA

The conventional GRAPPA formulation in Eqs. 1 or 2 models the calibration and reconstruction as a standard linear regression and prediction problem, where the undersampled part of the ACS corresponds to the regressors and the rest is the regressands. With this formulation, if the undersampled points of the ACS (regressors) are measured exactly or observed without noise, and noise is present only in the "missing" ones of the ACS (regressands), then the least-squares solution is optimal and the error in the reconstruction is proportional to the input noise. However, this is not the case in GRAPPA because all ACS data are obtained from measurement and thus contain the same level of noise.

To understand the effect of noise in both parts of the ACS data (regressors and regressands), we describe the regression and prediction process of GRAPPA using latent variables (28). Specifically, if $\mathbf{A}$ and $\mathbf{b}$ are observed variables that come from the ACS data with measurement noise, we assume that there exist some unobserved latent variables $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ representing the true, noise-free counterparts, whose true functional relationship is modeled as a linear function f. We thereby have

$$\begin{cases} \mathbf{A} = \tilde{\mathbf{A}} + \delta_{\mathbf{A}} \\ \mathbf{b} = \tilde{\mathbf{b}} + \delta_{\mathbf{b}} \\ f: \ \tilde{\mathbf{b}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}} \end{cases} \quad [5]$$

where $\delta_{\mathbf{A}}$ and $\delta_{\mathbf{b}}$ represent measurement noises that are present in the ACS data and assumed to be independent of the true value $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$, and $\tilde{\mathbf{x}}$ denotes the latent true coefficients for the linear relationship between $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ without the hidden noise.

In the standard regression process, the coefficients $\mathbf{x}$ is estimated by fitting to the observed data in $\mathbf{A}$ and $\mathbf{b}$:

$$\mathbf{b} = \mathbf{A}\mathbf{x} \quad \rightarrow \quad \tilde{\mathbf{b}} + \delta_{\mathbf{b}} = (\tilde{\mathbf{A}} + \delta_{\mathbf{A}})\mathbf{x}. \qquad [6]$$

Therefore, there is a bias $\delta_{\mathbf{x}} = \mathbf{x}\text{-}\tilde{\mathrm{x}}$ in the coefficients estimated from the least-squares fitting, where

$$\mathbf{x} = [(\tilde{\mathbf{A}} + \delta_{\mathbf{A}})^T(\tilde{\mathbf{A}} + \delta_{\mathbf{A}})]^{-1}(\tilde{\mathbf{A}} + \delta_{\mathbf{A}})^T(\tilde{\mathbf{b}} + \delta_{\mathbf{b}}). \qquad [7]$$

For example, consider the simplest case where $\mathbf{x}$ is a scalar and $\mathbf{b}$ and $\mathbf{A}$ are both column vectors whose elements b_t and a_t represent measurements at index t. The estimated coefficient is given by

$$\mathbf{x} = \sum_{t=1}^{T} a_t b_t \Big/ \sum_{t=1}^{T} a_t^2, \qquad [8]$$

which deviates from the true coefficient $\tilde{\mathrm{x}} = \tilde{\mathrm{b}}/\tilde{\mathrm{a}}$. When the number of measurements T increases without bound, the estimated coefficient converges to

$$x = \tilde{x}/(1 + \sigma_b^2/\sigma_A^2), \qquad [9]$$

where the noise in $\mathbf{A}$ and $\mathbf{b}$ is assumed to have zero mean and variance of σ_{A}^2 and σ_{b}^2, respectively. It suggests that even if there are an infinite number of measurements, there is still a bias in the least-squares estimator. Since the bias depends on the noise in both $\mathbf{A}$ and $\mathbf{b}$, its effects on the estimated coefficients $\mathbf{x}$ are also noise-like. In the multivariable case, the bias of GRAPPA coefficients is not easily characterized analytically, but is known to be upper bounded by (Theorem 2.3.8 in Ref. (29))

$$\frac{||\delta_x||_2}{||\mathbf{x}||_2} \leq \kappa(\tilde{\mathbf{A}})\left(\frac{||\delta_A||}{||\tilde{\mathbf{A}}||} + \frac{||\delta_b||}{||\mathbf{b}||} + \frac{||\delta_A||}{||\tilde{\mathbf{A}}||}\frac{||\delta_b||}{||\mathbf{b}||}\right), \qquad [10]$$

where $\kappa(\tilde{\mathrm{A}})$ is the condition number of matrix $\tilde{\mathrm{A}}$. The bound in Eq. 10 suggests that the bias in GRAPPA coefficients can be large at high reduction factors due to ill-conditioned $\mathbf{A}$ (12). In addition, the bias is not a linear function of the noise in the ACS. This is known as the errors-in-variable problem in regression. Figure 1 uses an example to demonstrate the nonlinearity of the bias for GRAPPA coefficients as a function of noise in the ACS data. Specifically, a set of brain data with simulated coil sensitivities (obtained from http://www.nmr.mgh.harvard.edu/~fhlin/) was used as the noise-free signal. We calculated the bias for GRAPPA coefficients (with the coefficients obtained from the noise-free signal as reference) when different levels of noise were added on all 24 lines of the ACS data. We plotted the normalized bias for GRAPPA coefficients as a function of the normalized noise level added to the ACS data. It is seen that the bias is not a linear function of noise level. However, when the noise is sufficiently low, the curve is well approximated by a straight line and the bias–noise relationship is approximately linear. Total least squares (30,31) is a linear method used to alleviate the problem by solving Eq. 6 using the total least squares instead of least squares. It addressed the error-in-variable problem to some extent when the noise is low. In the reconstruction

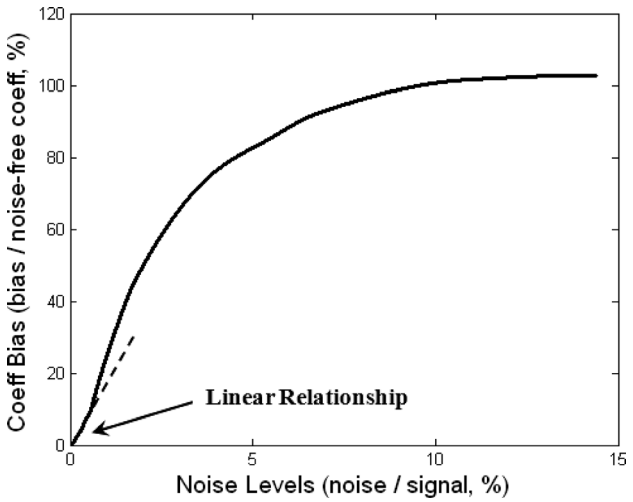

FIG. 1. Nonlinearity of the bias for GRAPPA coefficients as a function of noise in ACS data.

step of GRAPPA, when the biased coefficients $\mathbf{x}$ are applied upon the noisy undersampled data $\mathbf{A}$ to estimate the missing data in outer *k*-space, errors presented in the reconstruction are given by

$$\mathbf{y} - \tilde{\mathbf{A}}\tilde{\mathbf{x}} = (\tilde{\mathbf{A}} + \delta_{\mathbf{A}})(\tilde{\mathbf{x}} + \delta_{\mathbf{x}}) - \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{A}}\delta_{\mathbf{x}} + \tilde{\mathbf{x}}\delta_{\mathbf{A}} + \delta_{\mathbf{A}}\delta_{\mathbf{x}} \qquad [11]$$

It shows the effect of biased coefficients on the estimated missing *k*-space data is also nonlinear and noise-like. A comprehensive statistical analysis of noise in GRAPPA reconstruction can be found in (32).

Proposed Nonlinear GRAPPA

All existing GRAPPA derivatives are based on the linear model in Eq. 1 without considering the nonlinear bias due to noise in the ACS data. To address the nonlinear, noise-like errors in GRAPPA reconstruction, a kernel method is proposed to describe the nonlinear relationship between the acquired undersampled data and the missing data in the presence of noise-induced errors. Please note that the kernel used in this paper is different from the kernel usually used in GRAPPA literature to represent the *k*-space neighborhood for linear combination.

General Formulation Using Kernel Method

Kernel method (33–36) is an approach that is widely used in machine learning. It allows nonlinear algorithms through simple modifications from linear ones. The idea of kernel method is to transform the data nonlinearly to a higher dimensional space such that linear operations in the new space can represent a class of nonlinear operations in the original space. Specifically, given a linear algorithm, we map the data in the input space A to the feature space H via a nonlinear mapping $\Phi(\cdot)$: $A \rightarrow H$, and then run the algorithm on the vector representation $\Phi(\mathbf{a})$ of the data. However, the map may be of very high or even infinite dimensions and may also be hard to find. In this case, the kernel becomes useful to perform the

algorithms without explicitly computing $\Phi(\cdot)$. More precisely, a kernel is related to the mapping Φ in that

$$k(\mathbf{a}_1, \mathbf{a}_2) = < \Phi(\mathbf{a}_1), \Phi(\mathbf{a}_2) >, \quad \forall \mathbf{a}_1, \mathbf{a}_2 \in A, \qquad [12]$$

where $<,>$ represents the inner product. Many different types of kernels are known (36) and the most general used ones include polynomial kernel (37) and Gaussian kernel (38).

To introduce nonlinearity into GRAPPA, we apply a nonlinear mapping to the undersampled k-space data $\mathbf{a}_i = \{S_l(k_y + tR\Delta k_y, k_x + h\Delta k_x)\}$ in the neighborhood of each missing point where l counts all coils and t and h transverse the acquired neighboring k-space data in k_y and k_x directions, respectively. Under such a mapping, Eq. 2 is transformed to the following new linear system of $\mathbf{x}$:

$$\mathbf{b} = \Phi(\mathbf{A})\mathbf{x}, \qquad [13]$$

where $\Phi(\mathbf{A}) = [\Phi(\mathbf{a}_1), \Phi(\mathbf{a}_1), \ldots, \Phi(\mathbf{a}_M)]^T$, with $\mathbf{a}_i$ being the ith row vector of the matrix $\mathbf{A}$ defined in Eq. 2. The new matrix $\Phi(\mathbf{A})$ is of $M \times N_K$, where N_K is the dimension in the new feature space which is usually much higher than K. Equation 13 means the missing data in $\mathbf{b}$ is a linear combination of the new data in feature space which are generated from the original undersampled k-space data $\mathbf{A}$.

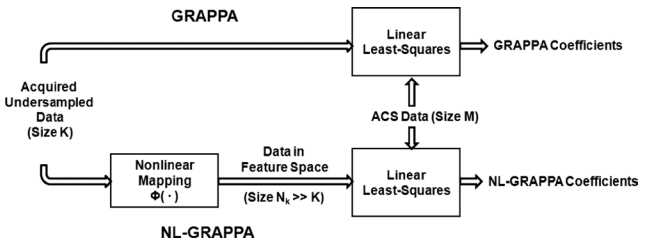

FIG. 2. Illustration of the calibration procedure for GRAPPA and nonlinear GRAPPA.

Although Eq. 13 is still a linear equation of the coefficients $\mathbf{x}$, it mathematically describes the nonlinear relationship between the undersampled and missing data because of the nonlinear mapping function $\Phi(\cdot)$. With the ACS data, the regression process to find the coefficients $\mathbf{x}$ in Eq. 13 for the proposed nonlinear GRAPPA can still be solved by a linear, least-squares algorithm in feature space

$$\hat{\mathbf{x}} = \left(\Phi^H(\mathbf{A})\Phi(\mathbf{A})\right)^{-1}\Phi^H(\mathbf{A})\mathbf{b}. \qquad [14]$$

Once the coefficients are estimated in Eq. 14, they are plugged back in Eq. 13 for the prediction process to reconstruct the missing data in outer k-space, like the

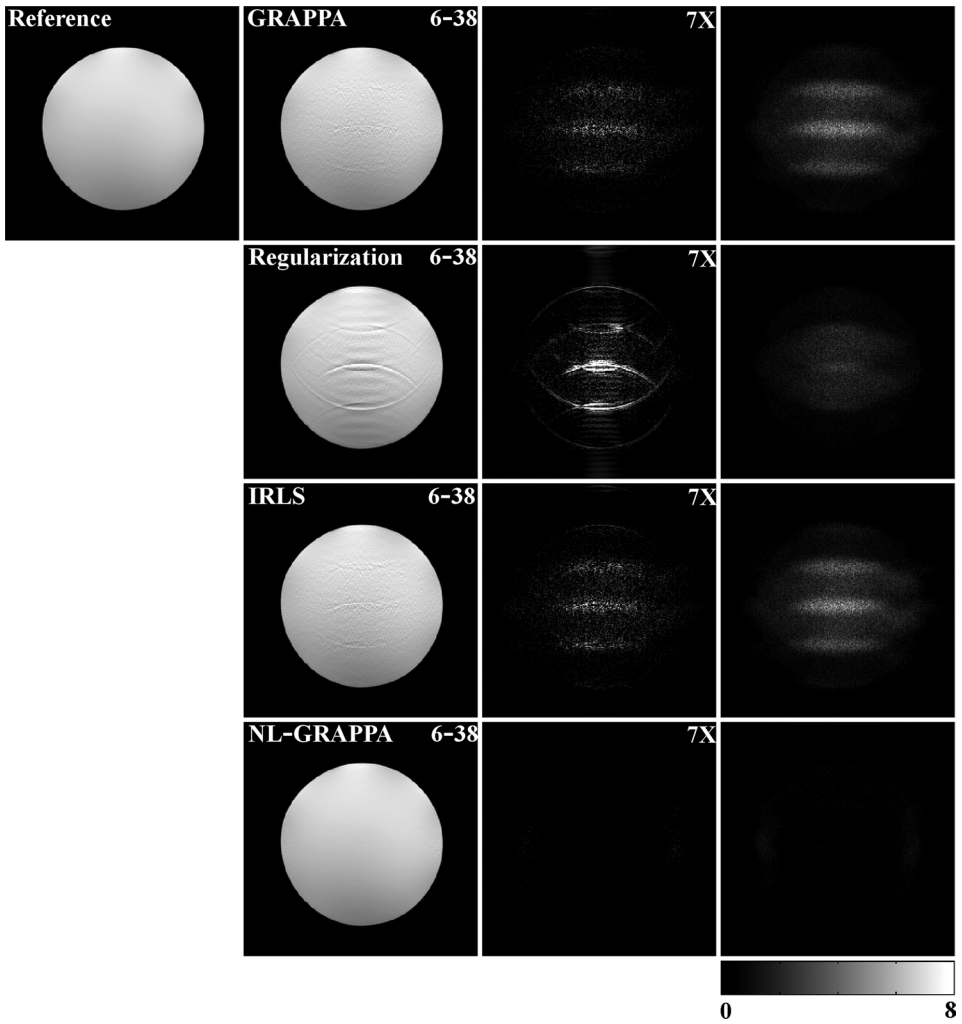

FIG. 3. Phantom images reconstructed from an eight-channel dataset with an ORF of 6 and 38 ACS lines (denoted as 6–38 on the right corner of each image). With the sum of squares reconstruction as the reference, the proposed nonlinear GRAPPA method is compared with conventional GRAPPA, regularized GRAPPA, and IRLS methods. The corresponding difference images with the reference (7× amplification) and g-factor maps are also shown on the right two columns, respectively.

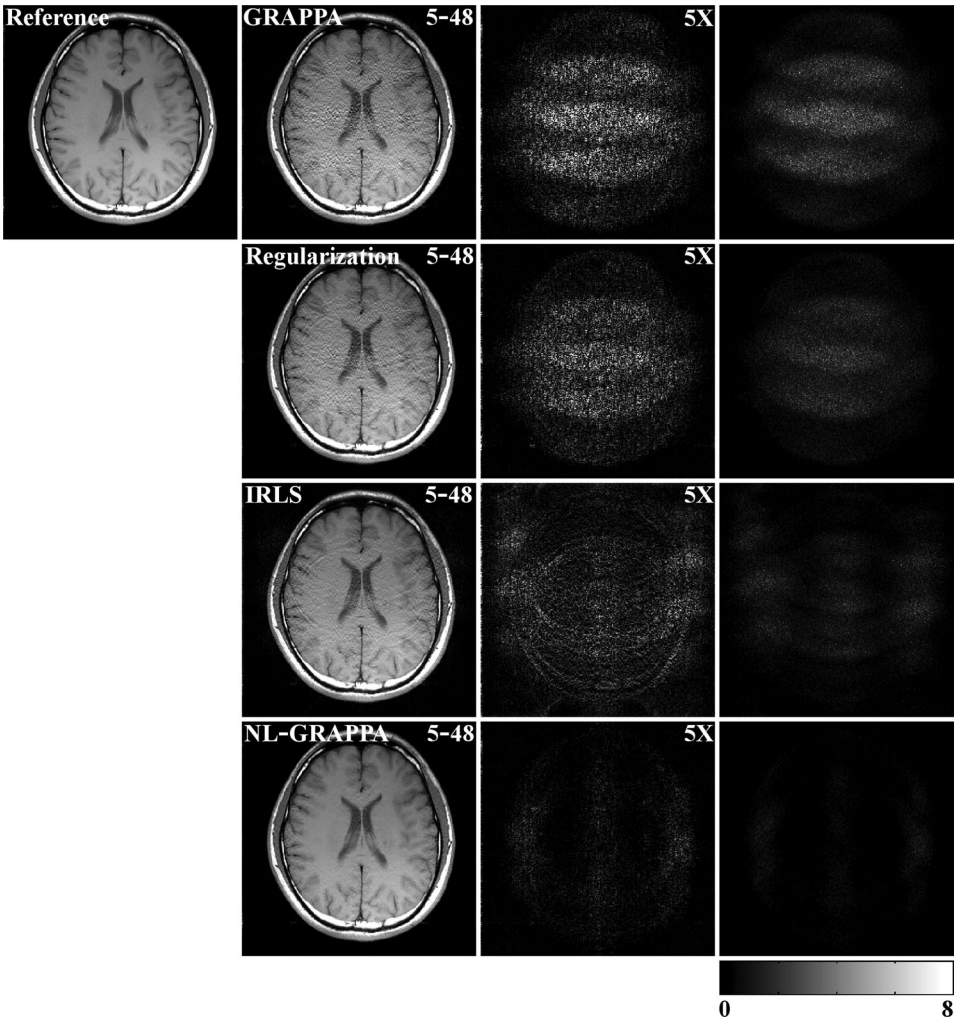

FIG. 4. Axial brain images reconstructed from a set of eight-channel data with an ORF of 5 and 48 ACS lines using GRAPPA, regularized GRAPPA, IRLS, and the proposed nonlinear method. The corresponding difference images with the reference (5× amplification) are shown on the middle column and *g*-factor maps on the right column.

conventional GRAPPA does. Figure 2 summarizes the above procedure and illustrates the nonlinear and linear parts of the proposed method (NL GRAPPA) in comparison to the conventional GRAPPA. It can be seen that the proposed method introduces an additional nonlinear mapping step into GRAPPA to pre-process the acquired undersampled data while the computational algorithm to find the coefficients is still the linear least-squares method. Other linear computational algorithms such as reweighted least-squares and total least-squares methods can also be used here.

Choice of Nonlinear Mapping $\Phi(\cdot)$

To choose the optimal kernel or feature space is not trivial. For example, Gaussian kernel has been proved to be universal, which means that linear combinations of the kernel can approximate any continuous function. However, overfitting of the calibration data may arise as a result of this powerful representation. Given the success of GRAPPA, we want the nonlinear mapping to be a smooth function that includes the linear one as a special case when the dimension of the feature space is as low as the original space. Since polynomials satisfy the desired properties, we choose an inhomogeneous polynomial kernel of the following form

$$\kappa(\mathbf{a}_i, \mathbf{a}_j) = \left(\gamma \mathbf{a}_i^T \mathbf{a}_j + r\right)^d, \quad [15]$$

where γ and r are scalars and d is the degree of the polynomial. Another advantage of polynomial kernel lies in the fact that its corresponding nonlinear mapping $\Phi(\mathbf{a})$ such that $k(\mathrm{a}_1, \mathrm{a}_2) = < \Phi(\mathrm{a}_1), \Phi(\mathrm{a}_2) >$ has explicit representations. For example, if $\gamma = r = 1$ and $d = 2$, $\Phi(\mathbf{a})$ is given by (39)

$$\Phi(\mathbf{a}) = \Big[1, \sqrt{2}a_1, \ldots, \sqrt{2}a_K, a_1^2, \ldots, a_K^2, \sqrt{2}a_1a_2, \ldots, \sqrt{2}a_ia_j, \ldots, \sqrt{2}a_{K-1}a_K\Big]^T, \quad [16]$$

where a_1, a_2, …, a_K are components of the vector $\mathbf{a}$ and there are $(K+2)(K+1)/2$ terms in total. It is seen that the vector includes the linear terms in the original space as well as the constant and second-order terms.

When all possible terms in $\Phi(\mathbf{a})$ are included, direct use of the kernel function may be preferred over the use of nonlinear mapping in Eq. 13 for the sake of computational complexity. However, our experiment (see Fig. 8) shows that the reconstruction using kernel functions suffers from blurring and aliasing artifacts. This is because the model is excessively complex and represents a too broad class of functions, and thus the model has been

overfit during calibration but poorly represents the missing data. This overfitting problem can be addressed by reducing the dimension of the feature space (40). The reduction of feature space is achieved here by keeping the constant term and all first-order terms $\sqrt{2}\mathrm{a}_1, \ldots, \sqrt{2}\mathrm{a}_k$, but truncating the second-order terms in vector $\Phi(\mathbf{a})$. Specifically, we sort the second-order terms according to the following order. We first have the square terms within each coil, and then the product terms between the nearest neighbors, the next-nearest neighbors, and so on so forth in *k*-space. The above order is then repeated for terms that are across different coils. With the sorted terms, we can truncate the vector $\Phi(\mathbf{a})$ according to the desired dimension of the feature space.

The performance of the proposed method depends on the number of second-order terms. If the number is too low, prediction is inaccurate because the feature space is not complex enough to accurately describe the true relationship between the calibration and undersampled data in presence of noise, and thus the reconstruction resembles GRAPPA and still suffers from noise-like errors. On the other hand, if the dimension is too high, the model is overfit by the calibration data but poorly represents the missing data, thus leading to aliasing artifacts and loss of resolution in reconstruction. This is known as the bias–variance tradeoff and is demonstrated using an example in Results section.

Explicit Implementation of Nonlinear GRAPPA

We find heuristically (elaborated in Results section) that it is sufficient to keep the number of the second-order terms to be about three times that of the first-order terms. That is, the feature space is reduced to

$$\tilde{\Phi}(\mathbf{a}) = [1, \sqrt{2}a_1, \sqrt{2}a_2, \cdots, \sqrt{2}a_K, \underbrace{a_1^2, a_2^2, \cdots, a_K^2}_{K}, \underbrace{a_1a_2, \cdots, a_ia_j, \cdots, a_{K-1}a_K}_{\sim K}, \underbrace{a_1a_3, \cdots, a_pa_q, \cdots, a_{K-2}a_K}_{\sim K}], \quad [17]$$

where (a_i, a_j) are nearest neighbors and (a_p, a_q) are next-nearest neighbors in *k*-space along k_x within each coil. We also find that a slight increase in the number of second-order terms does not change the reconstruction quality, but increases the computation.

After plugging the above truncated mapping vector $\tilde{\Phi}(\mathbf{a})$ in Eq. 17 into the matrix representation in Eq. 13 and changing to the notations in conventional GRAPPA, the proposed nonlinear GRAPPA method is thereby formulated as

$$\begin{aligned} S_j(k_y + r\Delta k_y, k_x) = {} & w_{j,r}^{(0)} \times 1 + \sum_{l=1}^{L}\sum_{b=B_1}^{B_2}\sum_{h=H_1}^{H_2} w_{j,r}^{(1)}(l,b,h) \times S_l(k_y + bR\Delta k_y, k_x + h\Delta k_x) \\ & + \sum_{l=1}^{L}\sum_{b=B_1}^{B_2}\sum_{h=H_1}^{H_2} w_{j,r}^{(2,0)}(l,b,h) \times S_l^2(k_y + bR\Delta k_y, k_x + h\Delta k_x) \\ & + \sum_{l=1}^{L}\sum_{b=B_1}^{B_2}\sum_{h=H_1}^{H_2-1} w_{j,r}^{(2,1)}(l,b,h) \times S_l(k_y + bR\Delta k_y, k_x + h\Delta k_x) \times S_l(k_y + bR\Delta k_y, k_x + (h+1)\Delta k_x) \\ & + \sum_{l=1}^{L}\sum_{b=B_1}^{B_2}\sum_{h=H_1}^{H_2-2} w_{j,r}^{(2,2)}(l,b,h) \times S_l(k_y + bR\Delta k_y, k_x + h\Delta k_x) \times S_l(k_y + bR\Delta k_y, k_x + (h+2)\Delta k_x), \end{aligned} \quad [18]$$

where the same notations are used as in Eq. 1.

The above nonlinear formulation represents a more general model for GRAPPA, which includes the conventional GRAPPA as a special case. It is seen that the first-order term of nonlinear GRAPPA in Eq. 18 is equivalent to the conventional GRAPPA, which mainly captures the linear relationship between the missing and acquired undersampled data in the absence of noise and approximations. The second-order terms of Eq. 18 can be used to characterize other nonlinear effects in practice such that noise and approximation errors are suppressed in reconstruction. The proposed formulation is a nonlinear model in the sense that nonlinear combination of acquired data contributes to estimation of missing *k*-space data. However the computational algorithm is still linear because the new system equation in Eq. 13 is still a linear function of the unknown coefficients and can still be solved by the linear least-squares method.

To better interpret the nonlinear GRAPPA method, we can consider the nonlinear terms as additional virtual channels as done in Ref. (21). For example, the first-order term in Eq. 18 represent a linear combination of *L* physical channels, while each second-order term represents a set of additional *L* virtual channels. Therefore, there are 4*L* channels in total when Eq. 18 is used. More second-order terms provide more virtual channels. It is worth noting that different from the true physical channel, there is no equivalent concept of coil sensitivities for the virtual channels. This is because the additional virtual channels are nonlinear function of the original physical channels. For example, the "square channel" takes the square of the *k*-space data point-by-point. In image domain, this is equivalent to the sensitivity-modulated image convolves with itself. Therefore, the resulting image cannot be represented as the product of the original image and another independent "sensitivity" function. Another point to be noted is that the virtual

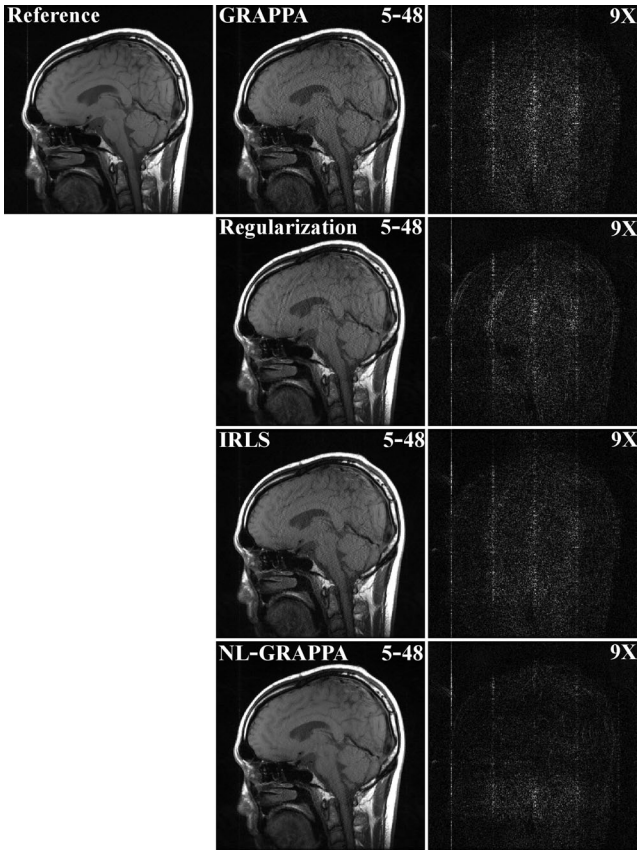

FIG. 5. Sagittal brain images reconstructed from a set of eight-channel data with an ORF 5 and 48 ACS lines and their corresponding difference images on the right. The proposed nonlinear GRAPPA suppresses most noise without aliasing artifacts.

channels are not necessarily all independent. Only adding channels that are linearly independent can improve the reconstruction performance. Choosing independent channels needs further study in our future work.

MATERIALS AND METHODS

The performance of the proposed method was validated using four scanned datasets. The first three scanned datasets were all acquired on a GE 3T scanner (GE Healthcare, Waukesha, WI) with an 8-channel head coil, and the last one was acquired on a Siemens 3T scanner (Siemens Trio, Erlangen, Germany). In the first dataset, a uniform water phantom was scanned using a gradient echo sequence (TE/TR = 10/100 ms, 31.25 kHz bandwidth, matrix size = 256 × 256, FOV = 250 mm^2). The second dataset was an axial brain image acquired using a 2D spin echo sequence (TE/TR = 11/700 ms, matrix size = 256 × 256, FOV = 220 mm^2). The third one was a sagittal brain dataset acquired using a 2D spin echo sequence (TR = 500 ms, TE = min full, matrix size = 256 × 256, FOV = 240 mm^2). In the fourth dataset, cardiac images were acquired using a 2D trueFISP sequence (TE/TR=1.87/29.9 ms, bandwidth 930 Hz/pixel, 50 degree flip angle, 6mm slice thickness, 34 cm FOV in readout direction, 256 × 216 acquisition matrix) with a 4-channel cardiac coil. Informed consents were obtained for all in vivo experiments in accordance with the institutional review board policy.

The proposed method was compared with conventional GRAPPA, as well as two existing methods that improve the SNR, Tikhonov regularization (12), and iterative reweighted least squares (IRLS) (14). The root sum of squares reconstruction from the fully sampled data of all channels was shown as the reference image for comparison. The size of the coefficients (blocks by columns) was chosen optimally for each individual method by comparing the mean-squared errors resulting from different sizes. The *g*-factor map was calculated using Monte Carlo simulations as described in (41) and used to show noise amplification. It is worth noting that for nonlinear algorithms, the SNR loss depends on the input noise level, and the *g*-factors shown in Results section are valid only in a small range around the noise level used in this study. Difference images were used to show all sources of errors, including blurring, aliasing, and noise. All methods were implemented in MATLAB (Mathworks, Natick, MA). To facilitate visual comparison, difference images from the reference and zoomed-in patches were also shown for some reconstructions. A software implementation of the proposed nonlinear GRAPPA method is available at https://pantherfile.uwm.edu/leiying/www/index_files/software.htm.

RESULTS

Phantom

Figure 3 shows the reconstructions of the phantom using sum of squares, GRAPPA, Tikhonov regularization, IRLS, and the proposed nonlinear GRAPPA for an ORF of 6 and the ACS of 38 (net acceleration of 3.41). The size of the coefficients was chosen optimally for each individual method, though the image quality is not sensitive to the change of size within a large range of the optimal choice. The size of the coefficients for nonlinear GRAPPA was two blocks and 15 columns and that for the other methods was four blocks and nine columns. It is seen that the conventional GRAPPA suffers from serious noise. Tikhonov regularization and IRLS can both improve the SNR to some extent but at the cost of aliasing artifacts. The proposed nonlinear GRAPPA method suppresses most of the noise without additional artifacts or loss of resolution. In addition, difference images with the reference and *g*-factor maps shown in Fig. 3 also suggest that the noise-like errors have quite different distributions spatially and they are more uniformly distributed in nonlinear GRAPPA than in other methods.

In Vivo Brain Imaging

Figures 4 and 5 show the reconstruction results for the two in vivo brain datasets, axial and sagittal, respectively. An ORF of 5 and the ACS of 48 were used with a net acceleration of 2.81. Nonlinear GRAPPA used a size of two blocks and 15 columns, while the other methods used that of four blocks and nine columns. The difference images with the reference are also shown (amplified five and nine times for display) in both Figs. 4 and 5 and *g*-factor maps are shown for the axial dataset in Fig. 4. It is seen that the reconstruction using the proposed

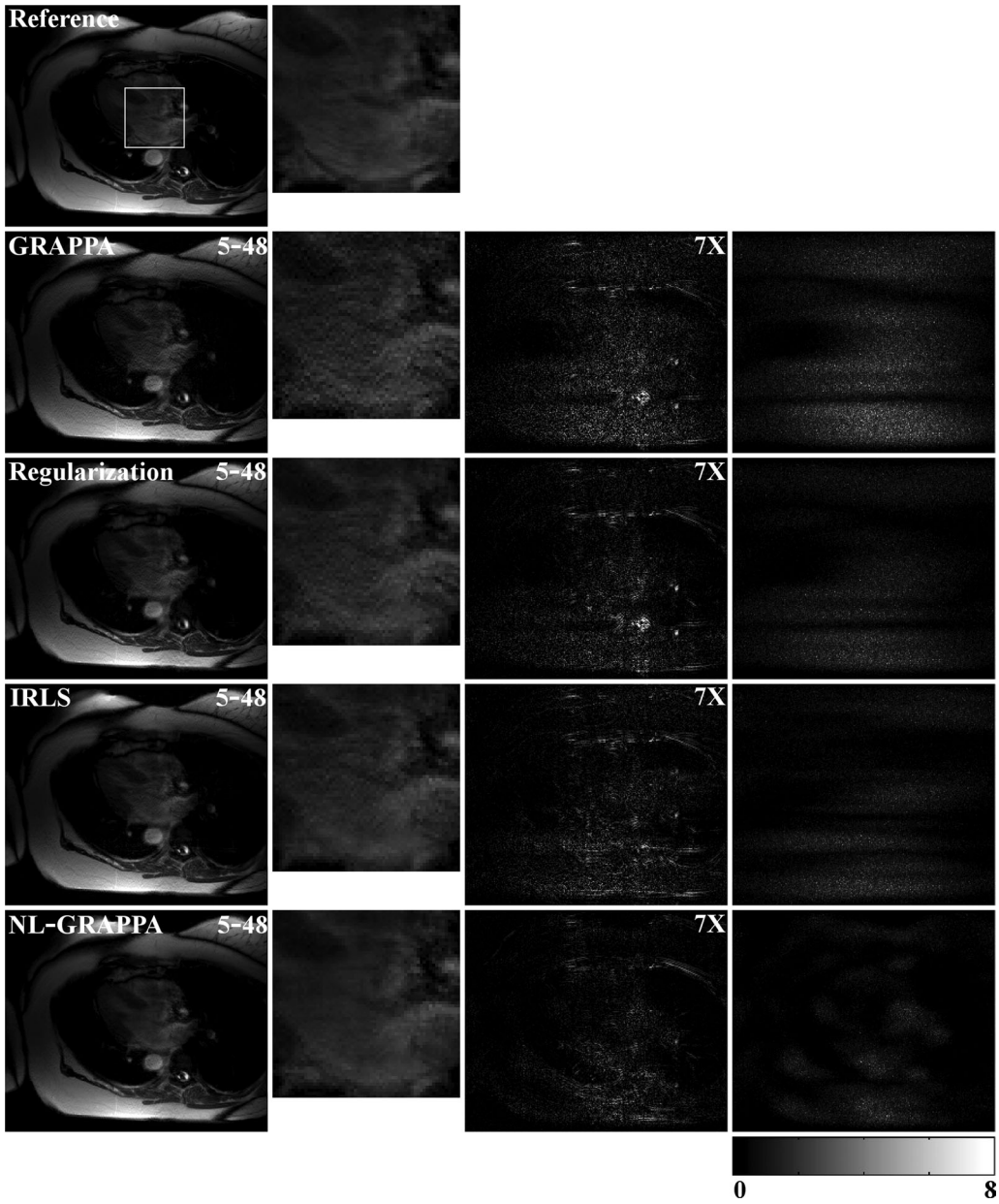

FIG. 6. Results from the four-channel cardiac dataset with an ORF 5 and 48 ACS lines. The reconstructed images, zoomed ROI, difference images, and *g*-factor maps are shown from left to right, respectively. They show that the proposed method can remove more noise than other methods while still preserving the resolution.

method achieves a quality superior to all other methods. The proposed method effectively removes the spatially varying noise in the conventional GRAPPA reconstruction without introducing aliasing artifacts as Tikhonov regularization and IRLS methods do. Furthermore, the proposed method also preserves the resolution of the axial image without blurring. There is only a slight loss of details in the sagittal image due to the tradeoff between noise suppression and resolution preservation (discussed later in Fig. 8).

In Vivo Cardiac Imaging

Figure 6 shows the results for the in vivo cardiac dataset in long axis. The ORF is 5 and number of ACS lines is 48 (net acceleration of 2.60). The size of the nonlinear GRAPPA coefficients was four blocks and 15 columns. The other methods used a size of four blocks and three columns. The ventricle areas are zoomed to show more details. Both the difference images and the *g*-factor maps are shown for all methods. The same conclusion can be made that the nonlinear GRAPPA method can significantly suppress the noise in GRAPPA and still preserve the resolution and avoid artifacts.

We also used the cardiac dataset to study how the number of second-order terms affects the nonlinear GRAPPA reconstruction quality. Specifically, we truncate all the sorted second-order terms to keep the number to be N times (e.g., three times in Eq. 18) that of the first-order terms. The normalized mean squared errors

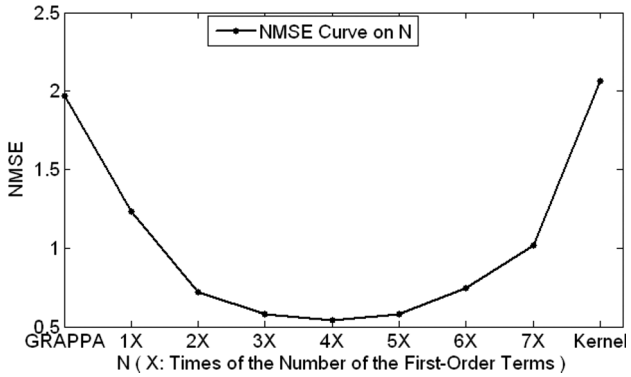

FIG. 7. Normalized mean squared error curves of the proposed method as a function of the number of the second-order terms using the long-axis cardiac dataset. The "U" shape of the curve suggests that some intermediate number should be chosen.

(NMSE) was calculated and plotted as the function of the number of the first-order terms in Fig. 7. In consideration of computational complexity, only the central 64 columns of the 48 ACS lines were used here for calibration. The two endpoints of the curve are the extreme cases of the proposed method. The left one corresponds to conventional GRAPPA without the second-order terms, and the right one is the case where all second-order terms are included (implemented efficiently using kernel representation directly). Figure 8 shows the corresponding reconstructions at some points of the curve. Both the curve and the images suggest that too small or too large N deteriorates reconstruction quality. When the number N increases, noise is gradually suppressed, but the resolution gradually degrades and aliasing artifacts gradually appear due to the overfitting problem. The optimal range for the value of N to balance the tradeoff between noise, resolution, and aliasing artifacts is seen to be 3-4 times of the number of the first-order terms, according to both the normalized mean squared error curve and the images. Because the value of N directly affects the computational complexity, $N = 3$ was chosen and shown to work well for all datasets tested in this study.

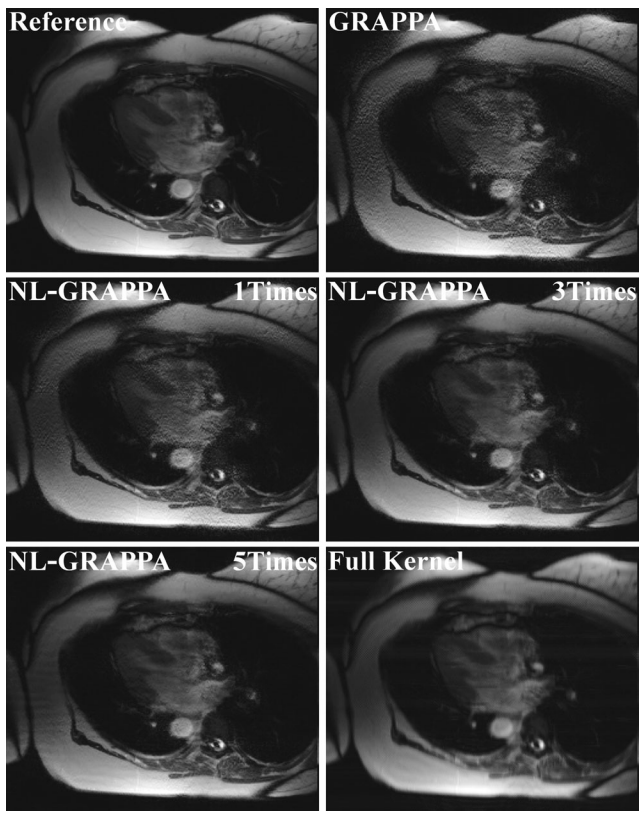

FIG. 8. The nonlinear GRAPPA reconstructions with an increasing number of the second-order terms show that the noise is gradually removed but artifacts gradually increase. In the extreme case, when all second-order terms are included, both blurring and aliasing artifacts are serious.

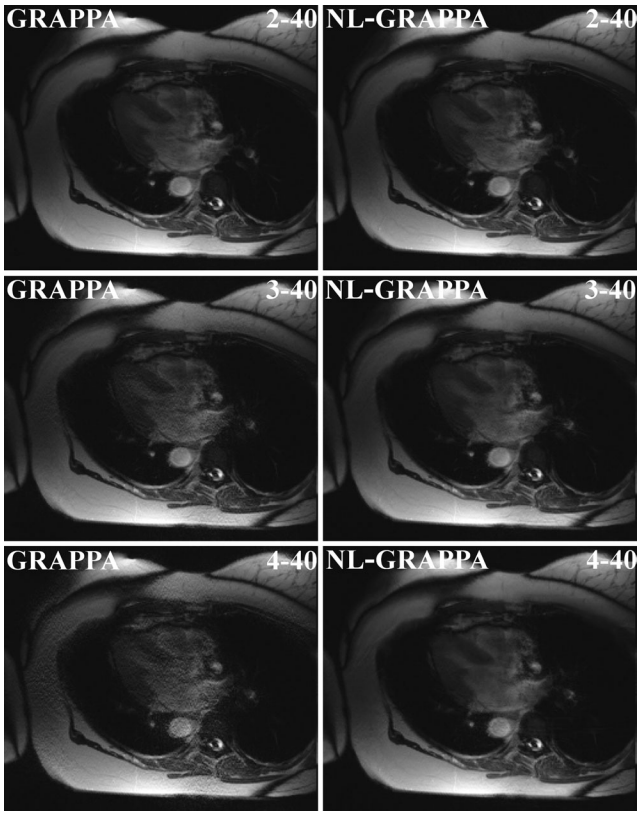

FIG. 9. Comparison between GRAPPA and nonlinear GRAPPA when ORF increases with fixed ACS lines. Contrary to GRAPPA, noise in the proposed method does not increase with the ORF.

DISCUSSION

We have shown in Results section that the proposed nonlinear GRAPPA method can outperform GRAPPA at high ORFs but also with a large number of ACS lines. It is interesting to see how the method behaves at lower ORFs or with fewer ACS lines. In Fig. 9, we compare GRAPPA and nonlinear GRAPPA with decreasing ORFs when the number of ACS lines is fixed to be 40. At a low ORF of 2, both methods perform similarly well. The proposed method has a slightly lower level of noise. As ORF increases, GRAPPA reconstruction begins to deteriorate due to the increased level of noise. In contrast, the

nonlinear GRAPPA method can maintain a similar SNR. Therefore, the benefit of nonlinear GRAPPA becomes more obvious at high ORFs. On the other hand, as ORF increases, the required number of ACS lines usually needs to increase to avoid aliasing artifacts.

Theoretically, the proposed method needs more ACS lines than GRAPPA to set up sufficient number of equations to avoid the aliasing artifacts. This is because there are more unknown coefficients to be solved for in the high dimensional feature space. Figure 10 shows how the reconstruction quality improves when increasing the number of ACS lines. The improvement of GRAPPA is primarily in terms of noise suppression while the improvement of nonlinear GRAPPA is in aliasing-artifacts reduction. Although more number of ACS lines is needed in the proposed method to avoid artifacts, the ORF can be pushed much higher than GRAPPA and thereby the net acceleration factor can remain low. For example, the combination of ORF of 5 and 48 ACS lines (denoted as 5–48) in Fig. 6 has a higher net acceleration factor to that of 4–40 (net acceleration of 2.57) in Fig. 9 and 3–32 (net acceleration of 2.30) in Fig. 10. The nonlinear GRAPPA reconstruction with 5–48 is always superior to GRAPPA with 5–48, 4–40, or 3–32 combinations.

In GRAPPA, it is known that the size of the coefficients also affects the reconstruction quality. More columns usually improves the data consistency and reduces aliasing artifacts, but at the cost of SNR and computation efficiency (42). A rather small number of columns (e.g., 3–5 columns) are typically used to balance the trade-off. In the proposed nonlinear GRAPPA, the size of coefficients also plays an important role. Figure 11 shows the GRAPPA and nonlinear GRAPPA reconstructions with 5, 9, and 15 columns of coefficients. Contrary to the observation in GRAPPA, more columns in nonlinear GRAPPA can improve the SNR due to the higher degree of freedom in calibration. In consideration of the computation cost, 15 columns were chosen to be used in our experiments.

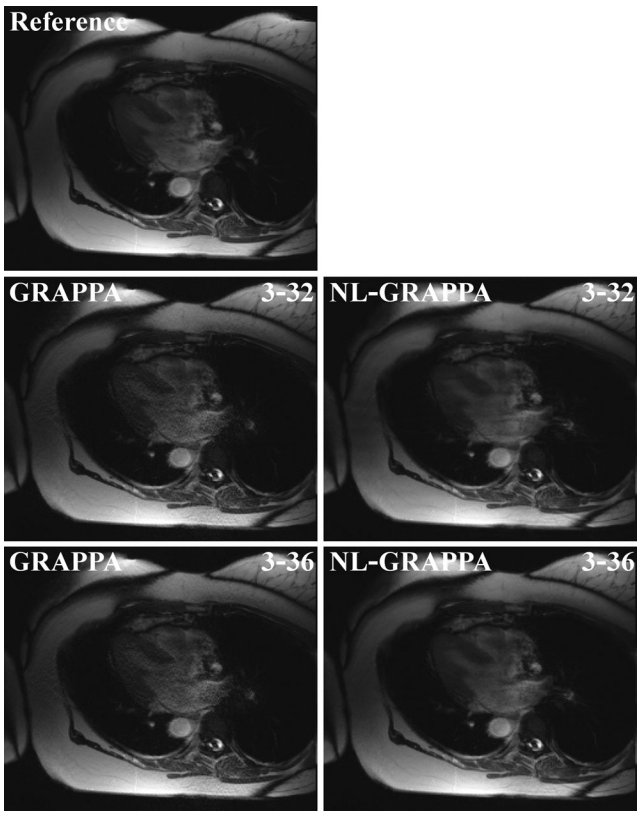

FIG. 10. Comparison of GRAPPA and nonlinear GRAPPA reconstructions when ACS increases with fixed ORF. It shows nonlinear GRAPPA needs more ACS lines than GRAPPA to avoid aliasing artifacts, but GRAPPA has more noise than nonlinear GRAPPA.

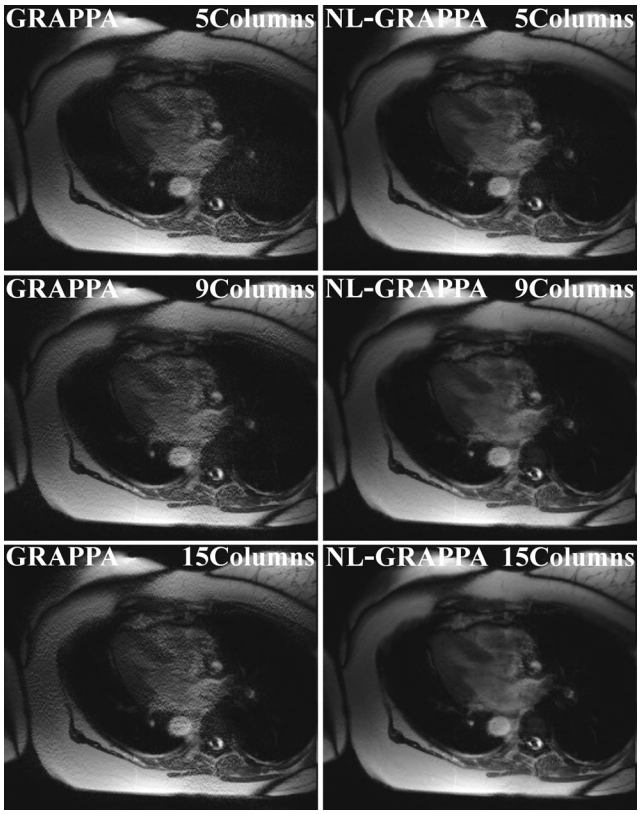

FIG. 11. Comparison between GRAPPA and nonlinear GRAPPA when different numbers of columns are chosen for the coefficients. Contrary to GRAPPA, the use of more columns in nonlinear GRAPPA can suppress more noise in reconstruction.

The computation time of the proposed method is about 2–5 times that of conventional GRAPPA and Tikhonov-regularized GRAPPA, while IRLS is the most time consuming among all. Furthermore, regularization can also be easily incorporated into the proposed reconstruction method.

CONCLUSION

In this paper, we propose a novel kernel-based nonlinear reconstruction algorithm for GRAPPA. The proposed method provides a more general model to characterize the noise behavior in GRAPPA reconstruction and thereby improves the SNR significantly. Experimental results demonstrate that the nonlinear GRAPPA is superior to conventional GRAPPA and some of the improved GRAPPA methods for the same net acceleration factor. Future work will investigate automatic reduction of

feature space using the methods in Refs. (43–45) to reduce the required number of ACS lines and improve the computational efficiency. We anticipate that the proposed nonlinear approach can bring further benefits to current applications of conventional GRAPPA.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation CBET-0731226, CBET-0846514, and a grant from the Lynde and Harry Bradley Foundation. The authors would like to thank Dr. Edward DiBella for providing the cardiac dataset.

REFERENCES

1. Griswold MA, Jakob PM, Heidemann RM, Mathias Nittka, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47:1202–1210.
2. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn Reson Med 1997;38:591–603.
3. Jakob PM, Griswold MA, Edelman RR, Sodickson DK. AutoSMASH: a self-calibrating technique for SMASH imaging. MAGMA 1998;7:42–54.
4. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. Magn Reson Med 1999;42:952–962.
5. Griswold MA, Jakob PM, Nittka M, Goldfarb JW, Haase A. Partially parallel imaging with localized sensitivities (PILS). Magn Reson Med 2000;44:602–609.
6. Kyriakos WE, Panych LP, Kacher DF, Westin CF, Bao SM, Mulkern RV, Jolesz FA. Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (SPACE RIP). Magn Reson Med 2000; 44:301–308.
7. Heidemann RM, Griswold MA, Haase A, Jakob PM. VD-AutoSMASH imaging. Magn Reson Med 2001;45:1066–1074.
8. Yeh EN, McKenzie CA, Ohliger MA, Sodickson DK. Parallel magnetic resonance imaging with adaptive radius in *k*-space (PARS): constrained image reconstruction using *k*-space locality in radiofrequency coil encoded data. Magn Reson Med 2005;53:1383–1392.
9. Ying L, Sheng J. Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). Magn Reson Med 2007;57:1196–1202.
10. Park J, Zhang Q, Jellus V, Simonetti O, Li D. Artifact and noise suppression in GRAPPA imaging using improved *k*-space coil calibration and variable density sampling. Magn Reson Med 2005;53:186–193.
11. Wang Z, Wang J, Detre JA. Improved data reconstruction method for GRAPPA. Magn Reson Med 2005;54:738–742.
12. Qu P, Wang C, Shen GX. Discrepancy-based adaptive regularization for GRAPPA reconstruction. J Magn Reson Imaging 2006;24:248–255.
13. Lin F-H, Prior-Regularized GRAPPA Reconstruction. Proceedings of the 14th Annual Meeting of ISMRM, Seattle, 2006. p. 3656.
14. Huo D, Wilson D.L. Robust GRAPPA reconstruction and its evaluation with the perceptual difference model. J Magn Reson Imaging 2008;27:1412–1420.
15. Huang F, Li Y, Vijayakumar S, Hertel S, Duensing GR. High-pass GRAPPA: a image support reduction technique for improved partially parallel imaging. Magn Reson Med 2008;59:642–649.
16. Nana R, Zhao T, Heberlein K, LaConte SM, Hu X. Cross-validation-based kernel support selection for improved GRAPPA reconstruction. Magn Reson Med 2008;59:819–825.
17. Nana R and Hu X. Data consistency criterion for selecting parameters for *k*-space-based reconstruction in parallel imaging. Magn Reson Imaging 2010;28, 119–128.
18. Zhao T, Hu X. Iterative GRAPPA (iGRAPPA) for improved parallel imaging reconstruction. Magn Reson Med 2008;59:903–907.
19. Griswold MA, Blaimer M, Breuer F, Heidemann RM, Mueller M, Jakob PM. Parallel magnetic resonance imaging using the GRAPPA operator formalism. Magn Reson Med 2005;54:1553–1556.
20. Bydder M, Jung Y. A nonlinear regularization strategy for GRAPPA calibration. Magn Reson Imaging 2009;27:137–141.
21. Blaimer M, Gutberlet M, Kellman P, Breuer FA, Kostler H, Griswold MA. Virtual coil concept for improved parallel MRI employing conjugate symmetric signals. Magn Reson Med 2009;61:93–102.
22. Honal M, Bauer S, Ludwig U, Leupold J. Increasing efficiency of parallel imaging for 2D multislice acquisitions. Magn Reson Med 2009; 61:1459–1470.
23. Chen Z, Zhang J, Yang R, Kellman P, Johnston LA, Egan GF. IIR GRAPPA for parallel MR image reconstruction. Magn Reson Med 2010;63:502–509.
24. Wang H, Liang D, King KF, Nagarsekar G, Chang Y, Ying L. Crossed-Sampled GRAPPA for parallel MRI. Proceedings of the IEEE Engineering in Medicine and Biology Conference, Buenos Aires, Argentina; 2010. p. 3325–3328.
25. Ying L, Abdelsalam E. Parallel MRI Reconstruction: A Filter-Bank Approach. Proceedings of the IEEE Engineering in Medicine and Biology Conference, Shanghai, China; 2005. p. 1374–1377.
26. Sharif B, Bresler Y. Distortion-Optimal Self-Calibrating Parallel MRI by Blind Interpolation in Subsampled Filter Banks. Proceedings of the IEEE International Symposium on Biomedical Imaging, Chicago, USA, 2011.
27. Huang F, Duensing GR. A Theoretical Analysis of Errors in GRAPPA. Proceedings of the 14th Annual Meeting of ISMRM, Seattle; 2006. p. 2468.
28. Carreira-Perpiñán MÁ. Continuous latent variable models for dimensionality reduction and sequential data reconstruction. PhD thesis, University of Sheffield, UK, 2001.
29. Watkins D. Fundamentals of matrix computations, 2nd ed. New York, USA: Wiley-Interscience; 2002.
30. Golub G. Some modified matrix eigenvalue problems. SIAM Rev 1973;15:318–344.
31. Golub G, Loan CV. An analysis of the total least squares problem. SIAM J Numer Anal 1980;17:883–893.
32. Aja-Fernández S, Tristán-Vega A, Hoge WS, Statistical noise analysis in GRAPPA using a parameterized noncentral Chi approximation model. Magn Reson Med 2011;65:1195–1206
33. Tomaso P, Steve S. The mathematics of learning: dealing with data. Notice Am Math Soc 2003;50:537–544.
34. Liu W, Pokharel PP, Principe JC. The kernel least-mean-square algorithm. IEEE Trans Signal Process 2008;56:543–554.
35. Camps-Valls G, Rojo-Alvarez JL, Martinez-Ramon M. Kernel methods in bioengineering, signal and image processing. Idea Group Publishing, London, 2007.
36. Schlkopf B, Smola AJ. Learning with kernels: support vector machinesregularizationoptimization and beyond (adaptive computation and machine learning). The MIT Press: Boston; 2001.
37. Frieb T, Harrison R. A Kernel-Based Adaline. Proceedings of the 7th European Symposium on Artificial Neural Networks, Bruges, Belgium; 1999. p. 245–250.
38. Kim K, Franz MO, Scholkopf B. Iterative kernel principal component analysis for image modeling. IEEE Trans Pattern Anal Machine Intell 2005;27:1351–1366.
39. Chang YW, Hsieh CJ, Chang KW, Ringgaard M, Lin CJ. Training and testing low-degree polynomial data mappings via linear SVM. J Machine Learn Res 2010;11:1471–1490.
40. Zhang T. Learning bounds for kernel regression using effective data dimensionality. Neural Comput 2005;17:2077–2098.
41. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, McKenzie CA. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and *k*-space-based parallel imaging reconstructions. Magn Reson Med 2008;60:895–907.
42. Bauer S, Markl M, Honal M, Jung BA. The effect of reconstruction and acquisition parameters for GRAPPA-based parallel imaging on the image quality. Magn Reson Med 2011; March, Early View Online.
43. Doneva M, Börnert P. Automatic coil selection for channel reduction in SENSE-based parallel imaging. MAGMA 2008;21(3):187–196.
44. Huang F, Vijayakumar S, Li Y, Hertel S, Duensing GR. A software channel compression technique for faster reconstruction with many channels. Magn Reson Imaging 2008;26:133–141.
45. Feng S, Zhu Y, Ji J. Efficient large-array k-domain parallel MRI using channel-by-channel array reduction. Magn Reson Imaging 2011;29: 209–215.