International Journal of Advances in Image Processing Techniques– IJIPT
Volume 1 : Issue 4 [ISSN 2372 – 3998]
Publication Date : 27 December, 2014

Limited memory quasi-Newton methods for nonlinear image restoration

Farzin Modarres Khiyabani

Abstract— **Variational models of unconstrained optimization problems *have been* extremely successful in a wide *variety* of *restoration problems*, such as image restoration. In this paper, we present an efficient limited memory quasi-Newton (QN) technique to compute meaningful solutions for large-scale problems arising in some image restoration problems. Numerical experiments reveals the effectiveness of the proposed method particularly for images of large size. *(Abstract)***

Keywords—Large-scale optimization, Image restoration, Quasi-Newton methods, Limited memory (key words)

I. Introduction

Restoring images from blurred or/and noisy data is an important task of image processing that is often formulated as an inverse problem to reconstruct the original image. The observation noise is usually assumed to be additive, white Gaussian that is independent of the image. However, in its most general form, image distortion is non-linear [1], and the noise can be either non-additive and/or non-Gaussian [7]. For example, the nonlinear behavior of image sensors becomes prominent when the light source changes rapidly during the exposure time [10]. The degradation could also be due to multiplicative noise [9]. Hunt [6] proposed expanding the observation model of the non-linear system into a Taylor series about the mean of the recorded image. Using this expansion an approximate filter is derived for image restoration. In this paper, we consider the following nonlinear space-invariant imaging system with additive noise

$$b = s(Au) + \eta,$$

where A is an $N \times N$ block Toeplitz, ill-conditioned matrix that characterizes blur, and b, u, and η are $N \times 1$ vectors. Here b is lexicographical ordering of the samples of the data (observed image); u is the true image (or object) and is nonnegative; η represents a particular realization of the additive (random) noise process that enters the data during the collection of the image.

Farzin Modarres Khiyabani
Department of Mathematics, East Azarbaijan Science and Research Branch, Islamic Azad University,
Tabriz, Iran.

We will consider the following nonlinear least squares problem with regularization

$$\min_u \quad Q(u) \equiv \min_u \| b - s(Au) \|_2^2 + \alpha \| u \|_2^2$$

to restore the original image. Here and throughout, $\|.\|$ denotes the usual Euclidean norm and α is a small positive number controlling the degree of regularity of the solution.

In order to optimize particular criteria, iterative algorithms can be designed to utilize general assumptions and require only deterministic parameters, such as the nonlinearity and the point spread function (psf) of the system. Thus, iterative algorithms are devoid of inaccuracies due to estimation of stochastic parameters from the data. In [11], Zervakis and Venetsanopoulos have considered the applications of the iterative Gauss-Newton (GN) method in nonlinear image restoration to solve the nonlinear least squares problem. However, the usage of Newton-type method requires calculations and storages of full matrices (in this context, the second derivative/Hessian matrix of $Q(u)$), which can be extremely expensive for many large-scale problems. Towards this direction, we consider limited memory quasi-Newton (QN) method based on the symmetric rank-one (SR1), which is arising from nonlinear image restoration problems. This limited-memory QN method resembles the QN method, except that the inverse Hessian approximation is defined implicitly as the outcome of updating a suitably selected initial matrix. In view of SR1 method's disadvantages and the nature of the image restoration problem (of large-scale and highly ill-conditioned), modifications on the limited memory scheme and SR1 updates are proposed to cater these difficulties.

SEEK DIGITAL LIBRARY

II. Limited Memory Quasi-Newton Methods

Consider the following unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable nonlinear function.

QN methods for unconstrained optimization of a nonlinear function are based upon the following iterative process

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k); \; k = 0,1,....$$

The famous SR1 update for the Hessian approximation can be derived by the formula

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k},$$

which makes a rank-one change to the previous Hessian approximation.

In 1992, Byrd *et al.* [5] presented the compact representations of the SR1 update. According to Byrd *et al.*, the limited memory SR1 formula can be written in a compact form:

$$B_k^{(L-SR1)} = B_0 + Q_k M_k^{-1} Q_k^T,$$

where $Q_k = Y_k - B_0 S_k$, $M_k = W_k - S_k^T B_0 S_k$,

$$S_k = [s_{k-m} \dots s_{k-2}\; s_{k-1}] \in \mathbb{R}^m, \quad Y_k = [y_{k-m} \dots y_{k-2}\; y_{k-1}] \in \mathbb{R}^m,$$

$$[W_k]_{ij} = \begin{cases} y_{i+k-m-1}^T s_{j+k-m-1}; & if \;\; i \le j, \\ y_{j+k-m-1}^T s_{i+k-m-1}; & if \;\; i > j. \end{cases}$$

Note that W_k is a symmetric matrix and its upper triangular is identical to that of $Y_k^T S_k$.

$$H_k^{(L-SR1)} = H_0 + (S_k - H_0 Y_k)(\hat{W}_k - Y_k^T H_0 Y_k)^{-1}(S_k - H_0 Y_k)^T,$$

where $\hat{W}_k$ is defined as the following symmetric matrix whose lower triangular is identical to that of $Y_k^T S_k$

$$[\hat{W}_k]_{ij} = \begin{cases} y_{j+k-m-1}^T s_{i+k-m-1}; & if \;\; i \le j, \\ y_{i+k-m-1}^T s_{j+k-m-1}; & if \;\; i > j. \end{cases}$$

However, like its counterpart in standard SR1 update, this compact representation may not preserve positive definiteness even when B_0 does.

A. *Modifications on Limited Memory SR1 Methods*

In this section we shall present several remedies to overcome the difficulties arise from the implementation of SR1 updates within the above mentioned limited memory schemes.

1) **Scaling for the Initial Matrix**

For standard QN methods, we often choose $B_0 = I$ initially, and after the first iteration, we use a multiple of identity, γI as initial matrix to update

$$B_1 = U(\gamma I, y_0, s_0),$$

where γ is usually chosen as the Oren-Luenberger scaling, given by $\gamma = \frac{y_0^T s_0}{s_0^T B_0 s_0} = \frac{y_0^T s_0}{s_0^T s_0}$.

We let $B_k^{(0)}$ be the scaling matrix $\gamma_k I$ where γ_k is computed by either

$$\gamma_k^{(1)} = \frac{\| y_{k-1}^T \|^2}{y_{k-1}^T s_{k-1}} \;\text{ or }\; \gamma_k^{(2)} = \frac{y_{k-1}^T s_{k-1}}{\| s_{k-1} \|^2},$$

as long as $y_{k-1}^T s_{k-1} > 0$.

2) **Small or Zero Denominator in SR1 Update**

It is known that the SR1 updating matrix $B_{k+1}^{(SR1)}$ is well defined only if both the following conditions hold for its denominator:

$$(y_k - B_k^{(SR1)} s_k)^T s_k \neq 0,$$

$$\cos\theta_k \equiv \frac{|(y_k - B_k^{(SR1)} s_k)^T s_k|}{\| y_k - B_k^{(SR1)} s_k \| \| s_k \|} \ge \rho > 0.$$

Although it rarely happens, once the cosine value approaches zero, it can hurt the performance of the algorithm. Thus, a safeguard that ensures

$$|(y_k - B_k^{(SR1)} s_k)^T s_k| \le \rho \| y_k - B_k^{(SR1)} s_k \| \| s_k \|,$$

is therefore strongly recommended.

3) **Positive Definiteness of SR1 Updating Matrix**

For L-SR1 method, we do not have any guarantee on the positive definiteness. Thus, we propose to use the Osborne and Sun's scaling selectively on certain iterations in preserving positive definiteness of SR1 update as follows: The detailed algorithm for computing ω is stated as below:

Algorithm 1.

IF b>a THEN $\omega_k := 1$
ELSE
 $\Delta := \sqrt{(c/b)^2 - c/a}$;
 $\theta_1 := c/b + \Delta$;
 $\theta_2 := c/b - \Delta$;
 DEFINE $\psi(\theta) = \frac{c - b\theta}{b\theta - a\theta^2}$;
 IF $\psi(\theta_1) \ge \psi(\theta_1)$ THEN
  $\omega_k := 1/\theta_1$;
 ELSE
  $\omega_k := 1/\theta_2$;
 END IF;
END IF;
RETURN.

SEEK DIGITAL LIBRARY

Algorithm 2 describes the detailed steps for computing γ_k that based on $\bar{\gamma}_k$ while Algorithm 3 gives the procedure for calculating γ_k using $\underline{\gamma}_k$.

Algorithm 2.
IF k=0 THEN $\gamma_k := 1$;
ELSE
$[\bar{C}_k, \bar{N}_k] := eigen-decomposition(D_k^{-1}\hat{W}_k D_k^{-1})$;
IF min-eigenvalue$(\bar{N}_k) > 0$ THEN
$\bar{\gamma}_k :=$ max-eigenvalue$(\bar{N}_k^{-\frac{1}{2}}\bar{C}_k^T D_k^{-1} Y_k^T Y_k D_k^{-1}\bar{C}_k \bar{N}_k^{-\frac{1}{2}})$;
$\gamma_k := \sigma_1 \times \bar{\gamma}_k$, $\sigma_1 = 1.1$;
ELSE IF $y_{k-1}^T s_{k-1} > 0$ THEN
$\gamma_k := \frac{\| y_{k-1} \|^2}{y_{k-1}^T s_{k-1}}$;
ELSE
$\gamma_k := \gamma_{k-1}$;
END IF;
END IF;
RETURN.

Algorithm 3.
IF k=0 THEN
$\gamma_k := 1$;
ELSE
$[C_k, N_k] :=$ eigen-decomposition$(D_k^{-1}W_k D_k^{-1})$;
IF min-eigenvalue$(N_k) > 0$ THEN
$\underline{\gamma}_k = 1/$max-eigenvalue$(N_k^{-\frac{1}{2}} C_k^T D_k^{-1} S_k^T S_k D_k^{-1} C_k N_k^{-\frac{1}{2}})$;
$\gamma_k := \sigma_2 \times \underline{\gamma}_k$, $\sigma_2 = 0.9$;
ELSE IF $y_{k-1}^T s_{k-1} > 0$ THEN
$\gamma_k := \frac{\| y_{k-1} \|^2}{y_{k-1}^T s_{k-1}}$;
ELSE
$\gamma_k := \gamma_{k-1}$;
END IF;
END IF;
RETURN.

Subsequently, we state the limited memory algorithm that uses these γ_k s (L-SR1-S):

Algorithm 4. (L-SR1-S Algorithm)
- Choose starting point x_0, integer $m > 0$, and an initial symmetric and positive definite starting matrix $H_0 = I$, set $k = 0$.
- If the convergence criterion is achieved, then stop.
- Compute $d_k = -H_k \nabla f(x_k)$.
- Find an acceptable steplength, α_k, such that the Wolfe conditions
$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k \nabla f(x_k)^T d_k$,
and
$\nabla f(x_k + \alpha_k d_k)^T d_k \ge \delta_2 \nabla^T f(x_k) d_k$,
for $0 < \delta_1 < \delta_2 < 1$, $\delta_1 < \frac{1}{2}$, are satisfied.
- Set $x_{k+1} = x_k + \alpha_k d_k$.
- Let $\tilde{m} = \min\{k, m-1\}$.

Compute γ_j via Algorithm 2 or 3. Update the scaled H_0, $\tilde{m}+1$ times i.e. let

$$\begin{aligned} H_{k+1} = {} & (W_k^T W_{k-1}^T .. W_{k-\hat{m}}^T)(\prod_{j=k-\hat{m}}^{k} \gamma_j H_0)(W_{k-\hat{m}} .. W_{k-1} W_k) \\ & + (W_k^T .. W_{k-\hat{m}+1}^T)(\prod_{j=k-\hat{m}+1}^{k} \gamma_j \mu_{k-\hat{m}} s_{k-\hat{m}} s_{k-\hat{m}}^T)(W_{k-\hat{m}+1} .. W_k) \\ & \vdots \\ & + W_k^T \gamma_k \mu_{k-1} s_{k-1} s_{k-1}^T W_k \\ & + \mu_k s_k s_k^T. \end{aligned}$$

- Set $k := k+1$ and go to Step 1.

III. Numerical Results

In this section, we employ our limited memory SR1 methods for solving large-scale optimization problems arising in image restoration. We have used the limited memory BFGS method with m=3,5 by Liu and Nocedal [4] and improved version of the conjugate gradient algorithm using the Fletcher–Reeve formula by Birigin and Martnez [3] (FRSCG), which is mainly a scaled variant of Perry's [8]. The gradient of the image function is given by
$\nabla_u Q(u) = -2A^T D_s(Au)[b - s(Au)] + 2\alpha u$, where $D_s(Au)$ denotes the diagonal matrix with the k-th diagonal entry being $[D_s(Au)]_{kk} = \frac{\partial_s}{\partial_x}(x)|_{x=\sigma_i A_{ki} u_i}$.

The Hessian matrix then can be written in the matrix representation: $A^T(D(Au))^2 A + \alpha I$.

We use the following point-wise nonlinear logarithmic function of Zervakis and Venetsanopoulos [11]

$$s(x) = 30 * log(x).$$

The discrete point spread function of the block-Toeplitz-Toeplitz-block matrix A is given by

$$a(x,y) = \exp[\frac{-(x^2+y^2)}{2}].$$

Noisy sequences are generated with signal-to-noise ratios (SNRs) of 30 and 40 dB, respectively. Observed images for noise-to-signal ratio of 40dB are shown in Figure 1. We employ the root-mean-square error (rmse) which describes the average relative deviation of the restored image $u(\alpha)$ from the original image u

$$rmse = \frac{\| u - u(\alpha) \|_2}{\| u \|_2},$$

where α is the the optimal regularization parameter. The parameters in Wolfe conditions are set as $\delta_1 = 10^{-4}$ and $\delta_2 = 0.9$ where $\alpha_k = 1$ is always tried first.

In the numerical tests, we consider the Cameraman test image from Matlab's Image Processing toolbox (the size of the images are 128×128). The number n of variables in the objective function is as large as or is equal to 16384. The parameter m is set to 3 and 5, respectively and the results are given in Table 1-4. The implementation of L-OCSSR1 uses the same line search technique as that of L-SR1 with compact representation. The only difference is in matrix representation and subsequently the step computation. In Table 1-4, the following notations are used:

- M1: L-OCSSR1 with $H_k^{(0)} = \|y_{k-1}\|^2 / y_{k-1}^T s_{k-1} I$;
- M2: L-SR1-S with $\gamma_k = 1.1 * \overline{\gamma}_k$;
- M3: L-SR1-S with $\gamma_k = 0.9 * \underline{\gamma}_k$;
- M4: FRSCG;
- M5: L-BFGS with $H_k^{(0)} = \|y_{k-1}\|^2 / y_{k-1}^T s_{k-1} I$.

Table 1-4 indicate that both L-SR1 (M2) and L-BFGS (M5) methods have some wins and losses in the relative errors and the number of iterations with L-SR1 (M2) scores most wins in average. Therefore, M2 and M5 seems to be more efficient than the other methods in terms of rmse and the iteration numbers. On the other hand, in term of CPU time, FRSCG (M4) wins definitely, since this method requires the least computational effort per iteration.

Table 1: The restoration results by Method 1-5 for the images of Bridge and Cameraman with 30 dB

Method	Bridge			Cameraman		
	rmse	itrn.	time	rmse	itrn.	time
M1 (m=3)	0.0750	219	311.2	0.0790	244	322.5
M2 (m=3)	0.0748	214	291.3	0.0789	238	301.4
M3 (m=3)	0.0754	243	335.1	0.0795	268	344.2
M4 (m=3)	0.0750	228	274.5	0.0791	255	386.1
M5 (m=3)	0.0749	216	288.4	0.0790	236	300.8

Table 2: The restoration results by Method 1-5 for the images of Bridge and Cameraman with 40 dB

Method	Bridge			Cameraman		
	rmse	itrn.	time	rmse	itrn.	time
M1 (m=3)	0.0768	261	336.8	0.0813	281	343.3
M2 (m=3)	0.0763	252	327.4	0.0807	270	334.2
M3 (m=3)	0.0772	292	341.7	0.0814	288	348.5
M4 (m=3)	0.0767	268	298.5	0.0821	290	319.3
M5 (m=3)	0.0761	255	321.8	0.0810	275	337.9

Table 3: The restoration results by Method 1-5 for the images of Bridge and Cameraman with 30 dB

Method	Bridge			Cameraman		
	rmse	itrn.	time	rmse	itrn.	time
M1 (m=3)	0.0619	139	217.3	0.0644	160	222.5
M2 (m=3)	0.0615	134	213.8	0.0647	157	218.4
M3 (m=3)	0.0631	150	233.6	0.0659	163	241.9
M4 (m=3)	0.0623	141	201.7	0.0662	178	208.5
M5 (m=3)	0.0613	138	215.2	0.0648	161	219.9

Table 4: The restoration results by Method 1-5 for the images of Bridge and Cameraman with 40 dB

Method	Bridge			Cameraman		
	rmse	itrn.	time	rmse	itrn.	time
M1 (m=5)	0.0648	175	286.9	0.0705	198	299.6
M2 (m=5)	0.0643	173	281.5	0.0695	192	290.9
M3 (m=5)	0.0653	190	302.1	0.0708	212	311.3
M4 (m=5)	0.06649	179	288.1	0.0719	197	286.4
M5 (m=5)	0.0646	176	283.5	0.0701	189	292.7

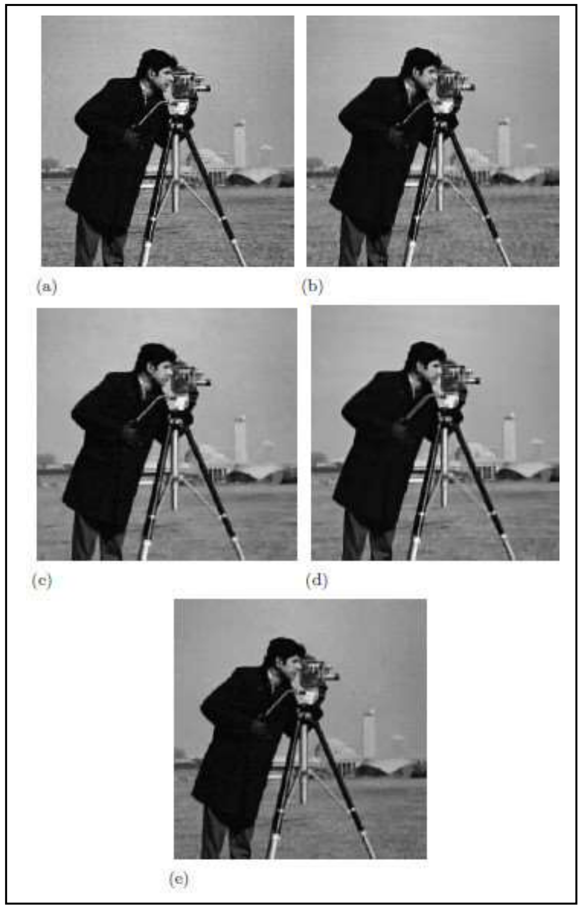

Figure 1: The restored images of Cameraman by using different methods (a): L-OCSSR1 with $H_k^{(0)} = (\|y_{k-1}\|^2 / y_{k-1}^T s_{k-1})I$, (b): L-SR1 with $H_k^{(0)} = (1.1 * \overline{\gamma}_k)I$, (c): L-SR1 with $H_k^{(0)} = (0.9 * \underline{\gamma}_k)I$, (d): FRSCG, and (e): L-BFGS with $H_k^{(0)} = (\|y_{k-1}\|^2 / y_{k-1}^T s_{k-1})I$.

SEEK DIGITAL LIBRARY

According to the results showed in Figures 1, we observe that images reconstructed by M1 and M2 are more clearly and better representing the original image. All in all, we can conclude that image reconstruction based on limited memory SR1 methods are useful.

IV. Conclusion

In summary, we have proposed some efficient limited memory symmetric rank-one methods for solving nonsmooth and nonseparable minimization problems. Theoretical considerations and numerical experiments indicate that our proposed limited memory SR1 method is competitive in solving the nonlinear image restoration problems and our numerical tests are able to demonstrate the efficiency of the proposed method. Furthermore, since the methods can also be served as general optimization solver, it may be fruitful to extend the methods for solving optimization problems related to image super-resolution and image denoising.

References

[1] H. C. Andrews and B. R. Hunt, Digital Image Restoration, Englewood Cliffs, NJ: Prentice-Hall, 1977.

[2] Y. Bard, "On a numerical instability of Davidon-like methods," Maths. Comput. vol. 22, pp. 665-666, 1968.

[3] E. Birgin and J. M. Martnez, "A spectral conjugate gradient method for unconstrained optimization," Appl. Math. Optim. vol. 43, pp. 117-128, 2001.

[4] R.H. Byrd, D.C. Liu, and J. Nocedal, "On the behaviour of Broyden's class of quasi-Newton methods," SIAM J. Optim. vol. 2, pp. 533-557, 1992.

[5] R.H. Byrd, J. Nocedal and R.B. Schnabel, Representations of Quasi-Newton Matrices and their use in Limited Memory Methods. Department of electrical engineering and computer science, Northwestern university, Technical Report NAM-03, 1992.

[6] B.R. Hunt, "Bayesian methods in nonlinear digital image restoration," IEEE Trans. Comput., vol. 26, no. 3, pp. 219-229, 1977.

[7] R.L. Lagendijk and J. Biemond, Iterative identification and restoration of imagess, Kluwer Academic Publishers, Boston, 1991.

[8] J.M. Perry, A class of conjugate gradient algorithm with a two step variable metric memory. Discussion Paper 269, Center for Mathematical studies in Economics and Management Science, Northwestern Univ, Chicago, 1977.

[9] A.M. Tekalp and G. Pavlovic, "Image restoration with multiplicative noise: Incorporating the sensor nonlinearity," IEEE Trans. Signal Process. vol. 39(9) pp. 2132-2136, 1991.

[10] He. Xiaofei, Y. Shuicheng and H. Yuxiao, "Face recognition using laplacianfaces," IEEE Trans. Pattern Anal. Machine lntell. vol. 27(3), pp. 328-340, 2005.

[11] M. Zervakis and A. Venetsanopoulos, "Iterative least squares estimators in nonlinear image restoration," IEEE Trans. Signal Process. vol. 40, pp. 927-945, 1992.

About Author:

Farzin Modarres Khiyabani received the B.S. and M.S. with a first class upper honors in 2005 and 2007. He furthered his study in UPM and obtained a degree of Ph.D in 2010 with specialization in Optimization. He received postdoctoral research associate offer from UPM in 2012. Since 2011, he has been a member of the faculty in the Mathematics Department at East Azarbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran. His research interests include scientific computing, nonlinear optimization, inverse problems, and image processing.

SEEK DIGITAL LIBRARY