Insight In to Bregman Algorithm

Arathy Reghukumar, Divya Haridas, SreekalaS ,K.P.Soman
Center for Excellence in Computational Engineering and Networking,
Amrita Vishwa Vidyapeetham,Amrita School of Engineering, Coimbatore -641112

ABSTRACT

Split-Bregman is a recent algorithm proposed with good convergence property in minimum number of iterations. It can be used in areas of denoising, deblurring, segmentation, inpainting etc. with ease due to its convergence property. In this paper we are trying to explore the fundamental theory of Bregman and Split Bregman with the help of convex function, constrained and unconstrained optimisation models.

Keywords: Convex function, Constrained and Unconstrained Optimisation, Bregman and Split-Bregman.

INTRODUCTION

Optimization plays an important role in image processing for example in case of denoising application, error between the original image and denoised image should be reduced. This minimisation process will be taken care in optimisation frame work. So bringing the filtering or denoising algorithm into this frame work gives an advantage when compared to conventional methods. Since the objective of operators like denoising, reconstruction etc are to bring out the appropriate image or by minimising error this is largely supported or achieved through convex optimization framework.

Consider a function F which represents an image processing problem like denoising ,restoration etc for a set of feasible solution can be found through optimization. Functions are of different types, widely used functions for optimisation are convex and concave because of its simplicity in finding the optimal minimum. Convex function is defined as if every line segment joining two points its graph is never below the graph.in order to find the solution for a convex/concave functions, constrained or unconstrained optimization can be used. In unconstrained optimisation search limit extends from $-\infty$ *to* ∞. But in constrained optimisation search limit is restricted, according to the subjected to condition.

CONVEX FUNCTION

A function can be defined as the relation between the set of input (domain) and set of output (codomain). Cube is an appropriate example for explaining convex set . All the points inside the cube constitute a convex set (S).

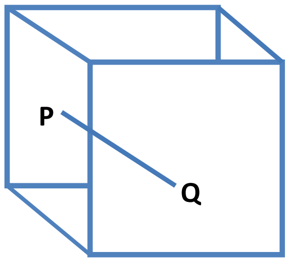

Figure:(1)

But in a non convex set, some of point lies outside the region R

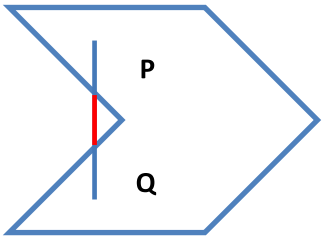

Figure:(2)

Consider the point R, which divides the line segment x1 and x2 joining the function which belongs to the convex set in the ratio of λ ,(1-λ) is shown in figure.

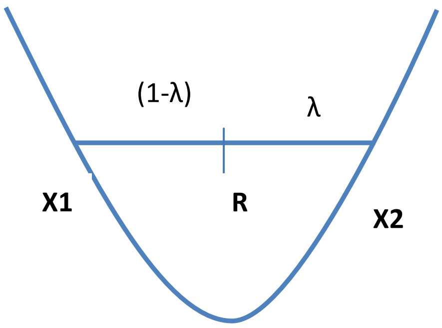

Figure:(3)

Point $R = \dfrac{\lambda x_1 + (1-\lambda)x_2}{\lambda + (1-\lambda)}$ (1)

$$R = \lambda x_1 + (1-\lambda)x_2 \quad (2)$$

is always lies in the convex set. This constitutes a function value. Function is said to be convex if and only if function F :X $\rightarrow R$ set X in a vector space ,for any two points x_1 and x_2 in X and $\lambda \in (0,1)$.

$$f(\lambda x_1 + (1-\lambda)x_2) \leq (\lambda f(x_1) + (1-\lambda)(fx_2)) \quad (3)$$

Since every point on a convex function are differentiable a convex function have unique gradient in every point except at minimum point.

$$\nabla f(x) = \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \frac{\partial f(x)}{\partial x_3} \ldots\ldots\ldots\ldots \frac{\partial f(x)}{\partial x_n} \quad (4)$$

Using Taylor series expansion the mathematical interpretation of convex function can be depicted as

$$f(x) = f(x_0 + (x - x_0) f'(x_0)) \quad (5)$$

For higher dimension

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) \quad (6)$$

The function $f(x)$ should be convex if satisfies the condition

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0) \quad (7)$$

CONSTRAINED AND UNCONSTRAINED OPTIMIZATION

Optimization problems are of two types –constrained and unconstrained optimization. Generally optimization is the method of finding maximum or minimum of a function. Consider a ‘n’variable function $f(x_1, x_2, \ldots.. x_n) = F(X)$ In order to find the solution of a function f we have to find a point X_0.

$$\begin{aligned} f(x_0) &\leq f(x) \\ f(x_0) &\geq f(x) \end{aligned} \quad (8)$$

The first inequality equation stands for minimizing the function and the latter equation stands for maximizing the function. For solving the unconstrained problems we use root finding algorithm .In unconstrained problems the search space is not related with any constraints. But in constrained optimization search space is related with equality and inequality constraints.An unconstrained minimization problem is

Minimize $f(x)$ (9)

Where the minimization is over all $x \in R^n$. A constrained optimization problem is defined as follows

Minimize $f(x)$ (10)

Subject to
$$\begin{aligned} g(x) &\leq 0 \\ h(x) &= 0 \end{aligned} \quad (11)$$

Where $f(x)$ is the objective function to be minimized, $g(x)$ is the set of inequality constraints and $h(x)$ is the set of equality constraints. Generally constrained optimization problems are Basis Pursuit Problem and TV Denoising Problem. To efficiently solve this kind of problems we use Bregman Iterative Algorithm. Basis Pursuit problem deals with finding the solution for linear systems of equations of the form $P\varphi = q$.

Consider the linear system of equations

$$\begin{aligned} &P_{11}\varphi_1 + P_{22}\varphi_2 + \ldots\ldots\ldots + P_{1n}\varphi_{1n} = q_1 \\ &P_{21}\varphi_1 + P_{22}\varphi_2 + \ldots\ldots\ldots. + P_{2n}\varphi_{2n} = q_2 \\ &. \\ &. \\ &P_{m1}\varphi_1 + P_{m2}\varphi_2 + \ldots\ldots\ldots. + P_{mn}\varphi_n = q_n \end{aligned} \quad (12)$$

$$P = \begin{bmatrix} P_{11} & P_{12} & \ldots\ldots P_{1n} \\ P_{21} & P_{22} & \ldots\ldots P_{2n} \\ \ldots\ldots\ldots\ldots\ldots & & \\ P_{m1} & P_{m2} & \ldots\ldots P_{mn} \end{bmatrix} \quad \varphi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \ldots \\ \varphi_n \end{bmatrix} \quad q = \begin{bmatrix} q_1 \\ q_2 \\ . \\ . \\ q_n \end{bmatrix}$$

The above system can be written in the form of ***where*** $P\varphi = q$, $P \in R^{mxn}$ $\varphi \in R^n$ and $q \in R^m$ .Here assume the rows of P are linearly independent .So the system of equation has infinite solution .Using minimal L_1 norm we can find the solution for linear system of equations and can be represented as

$$\min_{subject\ to\ P\varphi = q} \|\varphi\|_1 \quad (13)$$

Where $\|\varphi\|_1 = \sum_{i=1}^{n} |\varphi_i|$

There are many problems while solving the constrained equation (1), because of imposing constraints for search direction. So we go for unconstrained basis pursuit problem ie;

$$\min_{\mu \in R^n} \mu \| \varphi \|_2 + \frac{1}{2} \| P\varphi - q \|_2^2 \qquad (14)$$

Where "μ" is a positive constant. These types of problems are used for Compressive Sensing. The second type of problem is total variation denoising. The removal noise from an original image is called denoising. The problem is of the form

$$\min_{\varphi \in BV(\Omega)} \| \varphi \|_{BV} + X(\varphi) \qquad (15)$$

Where $\| \varphi \|_{BV}$ is not a smooth convex function and it is added with strictly convex function $X(\varphi)$. BV is the bounded variation of the form $\Omega \in R^n$ .To find the BV norm of a function has high computation cost and can be replaced by L_1 norm of the gradient ie; $\| \varphi \|_{BV} \to \| \nabla \varphi \|_1$ and $X(\varphi)$ can be replaced by

$$\min_{\varphi \in BV(\Omega)} \| \nabla \varphi \| + \frac{\mu}{2} \| \varphi - q \|_2 \qquad (16)$$

Here μ should be greater than zero.

In older days we were using interior point method. Consider a medium sized problem, using interior point algorithm requires upto 100 iteration for computing the solution with a relative tolerance of about 0.01 . So we go for Bregman Iterative Algorithm.

BREGMAN METHOD

Bregman is a highly specific and efficient algorithm for deblurring, denoising, segmentation etc. To solve general L_1 regularization problem use the equation

$$\arg\min_{u} | d(\varphi) | + \| P\varphi - q \|_2 \qquad \textbf{(17)}$$

To split the $L_1 and L_2$ component, introduce a new term to solve the constrain problem

$$\min_{\varphi, d} \| d \|_1 + X(\varphi) \qquad (18)$$

$$\min_{\mu \in R^n} \mu \| \varphi \|_2 + \frac{1}{2} \| P\varphi - q \|_2 \qquad (19)$$

Add a penalty term to make the constrained problem to unconstrained problem. By combining both (18) and (19) we get

$$\arg\min_{u,d} \| d \|_1 + X(\varphi) + \frac{\lambda}{2} \| d - d(\varphi) \|_2 \qquad (20)$$

Let $\| d \|_1 + X(\varphi)$ be $E(\varphi, d)$

then the equation becomes

$$\arg\min_{u,d} E(\varphi, d) + \frac{\lambda}{2} \| d - d(\varphi) \|_2 \qquad (21).$$

Now we can define the Bregman Distance

Bregman Distances

For a smooth differentiable convex function, gradient is possible . But for a non-differential function the concept sub-gradient comes into play. At a point A there can be many P, where P is the sub gradient. Set of all sub-gradients form Sub differential.

$$p \in \partial d(\varphi^k) . \qquad (22)$$

It forms a convex set

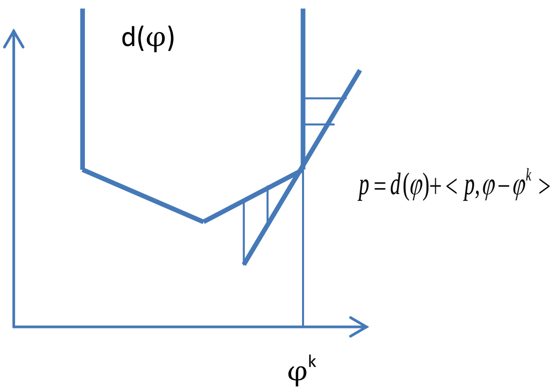

Figure:(4)

$$d(\varphi) \geq d(\varphi^k) + \langle p^k, \varphi - \varphi^k \rangle \quad (23)$$

P is the sub-gradient at $\varphi = \varphi^k$

In multi-dimensional,

$$d(\varphi) = d(\varphi^k) + \langle p, \varphi - \varphi^k \rangle \quad (24)$$

BregmanDistanceFunction

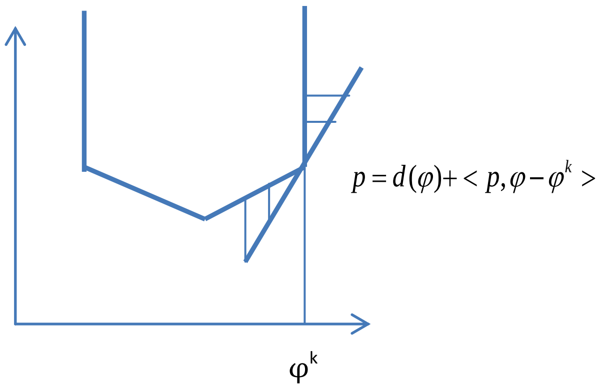

Figure: (5)

The function at P can be plotted as in figure

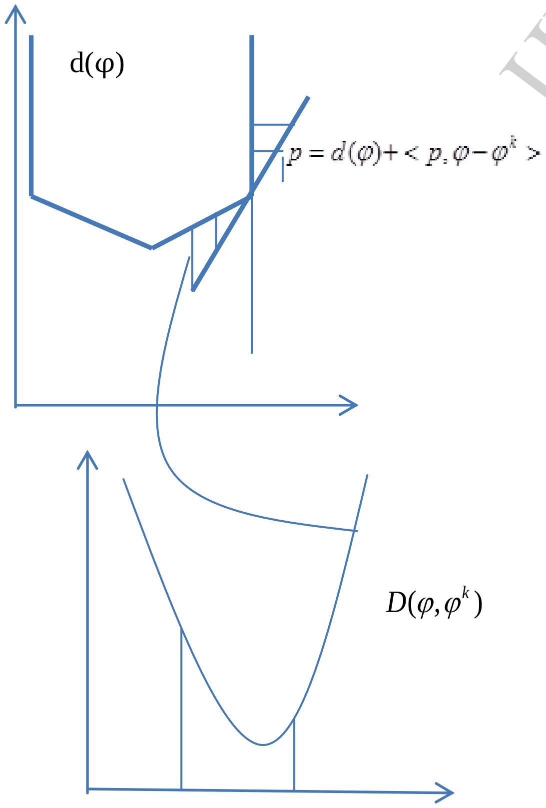

Figure:(6)

Where $d(\varphi^k) - \langle p^k, \varphi - \varphi^k \rangle$ is the equation of the sub gradient at φ^k

The distance is always positive. At $\varphi = \varphi^k$, $D_j^{p^k} = 0$ as seen from the graph.

Similarly for other u values $D_j^{p^k} \geq 0$

Our aim is to minimize

$$\min_{\varphi, d} \| d \|_1 + X(\varphi) \quad (25)$$

$$B_k(\varphi) = D_d^{p^{k-1}}(\varphi, \varphi^{k-1}) + X(\varphi) \quad (26)$$

Such that replace the semi convex function d(φ) with Bregman distance $D_d^{p^k}(\varphi, \varphi^k)$. We make an argument that $B_k(\varphi)$ is greater than zero.

$D_d^{p^{k-1}}(\varphi, \varphi^{k-1})$ is a distance function with minimum zero .This is strictly convex function. Therefore

$$\varphi^{k+1} = \arg\min_{\varphi} D_d^{p^k}(\varphi, \varphi^k) + X(\varphi^k) \quad (27)$$

At φ^k one of the sub-gradient is zero, $0 \in \partial d(\varphi^k)$ then,

$$B_k(\varphi) = d(\varphi) - d(\varphi^{k-1}) - < p^{k-1}, \varphi - \varphi^{k-1} > + X(\varphi) \geq 0 \quad (28)$$

$B_k(\varphi)$ is differentiating with respect to φ^k

$$0 \in \partial d(\varphi^k) - p^{k-1} + \nabla X(\varphi^k) \quad (29)$$

This (29) is not a regular function and to make it a regular function

$$0 = \partial d(\varphi^k) - p^{k-1} + \nabla X(\varphi^k) \quad (30)$$

Now it is easy to find the minimum value.

$$p^{k-1} - \nabla X(\varphi^k) \in \partial d(\varphi^k) \quad (31)$$

The gradient at k is

$$p^k = p^{k-1} - \nabla X(\varphi^k) \quad (32)$$

The gradient at k+1 is obtained by subtracting the derivative of strictly convex function at k+1 from previous gradient p at k.

$$p^{k+1} = p^k - \nabla X(\varphi^{k+1}) \qquad (33)$$

Bregman Algorithm

K=0, φ^0 , p^0=0

While φ^k not converge

$$\varphi^{k+1} = \arg\min D_j^{p^k}(\varphi, \varphi^{k+1}) + X(\varphi)$$

$$p^{k+1} = p^k - \nabla X(\varphi^{k+1}) \in \partial d(\varphi^{k+1})$$

$$k = k+1$$

end while

φ, p are the variables used in Bregman algorithm. K is the number of iteration. Successive iterations are given as φ^{k+1}, p^{k+1}. The algorithm runs in a while loop k is incrementing for successive iterations and it ends when φ^k converges.

SPLIT- BREGMAN ALGORITHM

Split Bregman algorithm is a suitable technique in solving convex minimisation problems which are of non- differentiable in nature. Optimisation problems of following format can be solved by using the split Bregman algorithm $d : \Omega \to R$ $X : \Omega \to R$ Constrained functioncan be defined as,

$$\min_{\varphi \in \Omega} \left\| d(\varphi) \right\|_1 + X(\varphi) \qquad (34)$$

Replace $d(\varphi)$ by $P\varphi$ therefore the equation can be written as

$$\min_{\varphi} \| P\varphi \|_1 + X(\varphi) \qquad (35)$$

Now take $d = P\varphi$

Therefore the objective function

$$\min_{\varphi, d} \quad \| d \|_1 + X(\varphi)$$

$subject\ to \quad d = P\varphi$ where φ, d s

Applying Lagrangian multiplier in above equation we can write

$$\min_{\varphi, d} \| d \|_1 + X(\varphi) + \frac{\lambda}{2} \| d - P\varphi \|_2 \qquad (36)$$

Now take $J(\varphi, d) = \| d \|_1 + X(\varphi)$

Now (36) can be written as

$$\min_{\varphi, d} J(\varphi, d) + \frac{\lambda}{2} \| d - P\varphi \|_2 \qquad (37)$$

As in iterative Bregman replace J(φ ,d) by $D_j^{p^k}(\varphi, \varphi^k, d, d^k)$ Therefore equation (37) can be rewritten as

$$\min_{\varphi, d} D_j^{p^k}(\varphi, \varphi^k, d, d^k) + \frac{\lambda}{2} \| d - P\varphi \|_2 \qquad (38)$$

(38) is in the form of

$$\min_{\varphi, d} \ J(\varphi, d) + X(\varphi, d) \qquad (39)$$

While updating to the next iterative point

$$(\varphi^{k+1}, d^{k+1}) = \min_{\varphi, d} D_j^{p^k}(\varphi, \varphi^k, d, d^k) + \frac{\lambda}{2} \| d - P\varphi \|_2 \qquad (40)$$

From Bregman Iterative Algorithm

$$p^{k+1} = p^k - \nabla X(\varphi^{k+1})$$

$$(\varphi^{k+1}, d^{k+1}) = \min_{\varphi, d} J(\varphi, d) + \lambda < b^k, d - P\varphi^k >$$

$$- + \frac{\lambda}{2} \| d - P\varphi^k \|_2$$

$$p_\varphi^{k+1} = p_\varphi^{k} - \lambda P^T (P\varphi^{k+1} - d^{k-1})$$

The p_φ^{k+1} is the gradient at k+1 with variable φ

Similarly

$$p_d^{k+1} = p_d^{k} - \frac{\partial}{\partial d} \left(\frac{\lambda}{2} \| d^{k+1} - P\varphi^{k+1} \|_2 \right) \qquad (41)$$

The $P_d^{\,k+1}$ is the gradient at k+1 with variable d.

$$p_d^{\,k+1} = p_d^{\,k} - \lambda(d^{k+1} - P\varphi^{k+1}) \quad (42)$$

$$p_d^{\,k+1} = p_d^{\,k} + \lambda(P\varphi^{k+1} - d^{k+1}) \quad (43)$$

Taking each term separately

$$= p_\varphi^{\,k} - \lambda P^T(-d^{k+1})$$
$$= P\varphi^{k-1} - \lambda P^T(P\varphi^k - d^k).....$$

and so on

$$p_\varphi^{\,k-1} = -\lambda P^T \sum_{i=1}^{k+1} P\varphi^i - d^i = -\lambda P^T b^{k+1} \quad (44)$$

Since $\sum_{i=1}^{k+1} P\varphi^i - d^i = b^{k+1}$ is called residual vector.

Similarly $p_d^{\,k-1} = \lambda \sum_{i=1}^{k+1} P\varphi^i - d^i = \lambda b^{k+1}$

Then,

$$(\varphi^{k+1}, d^{k+1}) = \min_{\varphi,d} J(\varphi,d) - J(\varphi^k,d^k) - < p_\varphi^{\,k}, \varphi - \varphi^k > - < p_d^{\,k}, d - d^k > + \frac{\lambda}{2}\| d - P\varphi \|_2 \quad (45)$$

Substitute $p_\varphi^{\,k} = -\lambda P^T b^k$ and $p^k_{\ \varphi} = \lambda b^k$ in (45), we get

$$(\varphi^{k+1}, d^{k+1}) = \min_{\varphi,d} J(\varphi,d) - J(\varphi^k,d^k) + \lambda < b^k, d - P\varphi^k > - \lambda < b^k, d - d^k > + \frac{\lambda}{2}\| d - P\varphi^k \|_2 \quad (46)$$

Since $< P^T b^k, \varphi - \varphi^k > = (P^T b^k)^T(\varphi - \varphi^k)$

$$= (b^k)^T P(\varphi - \varphi^k)$$

$$= < b^k, P(\varphi - \varphi^k) >$$

$$= < b^k, P\varphi - P\varphi^k >$$

$$= < b^k, d - P\varphi^k > \quad (47)$$

Then,

$$\min_{\varphi,d} J(\varphi,d) - \lambda(b^k, d - P\varphi) + \frac{\lambda}{2}\| d - P\varphi \|_2 + c \quad (48)$$

Here c= $-(J(\varphi^k,d^k) + \lambda < b^k, d - d^k >)$ T Therefore the equation can be rewritten as

$$\min_{\varphi,d} J(\varphi,d) + \frac{\lambda}{2}\| d - P\varphi - b^k \| + c_2 \quad (49)$$

Solving the above equation we get

$$b^{k+1} = b^k + (P\varphi^{k+1} - d^{k+1}) \quad (50)$$

Split-Bregman Algorithm

$$Intialize\ k = 0\ \varphi^0 = 0\ b^0 = 0$$
$$while \| \varphi^k - \varphi^{k-1} \|_2^2 > tol\ do$$
$$\varphi^{k+1} = \min_{\varphi} X(\varphi) + \frac{\lambda}{2}\| d^k - \phi(\varphi) - b^k \|_2^2$$
$$d^{k+1} = \min_{d} |d| + \frac{\lambda}{2}\| d - P(\varphi^{k+1}) - b^k \|_2^2$$
$$b^{k+1} = b^k + (P\varphi^{k+1} - b^k)$$
$$k = k+1$$
$$end\ while$$

φ, b, d are the variablesin Split-Bregman. In addition to the variables used in Bregman an extra variable is introduced here for reducing the computational complexity. K is the present iteration, and successive iteration are given as k+1. The difference in the value of present and previous iteration is compared with the tolerance value, if it is greater than the tolerance value then the while loop continues, then while loop ends.

EXPERIMENT AND RESULTS

The experimental part of Split- Bregman denoising is implemented using MATLAB. The denoised output of a cameraman image which has been corrupted with some random noiseis given below. The corresponding PSNR of 28.3971 and MSE of 94.0525 is obtained by using weighting parameter for fidelity term mu is .050 and tolerance is 0.001.

Figure (6): original

Figure (7): noisy

Figure (8): Denoised

Figure (9):Difference

CONCLUSION

In this paper we are trying to explore the fundamental theories and mathematical explanation of Bregman and Split-Bregman with the help convex function and constrained optimisation. After that explaining the denoising property of Split-Bregman using corresponding algorithms and the regarding results are depicted in above figure (6,7,8,9).

REFERENCES

1) Bregman Algorithms Author :JacquelineBush Supervisor Dr.Carlos Garc´ıa-Cervera June 10, 2011

2) ’’Rudin{Osher{Fatemi “Total Variation Denoising using Split Bregman’’
Pascal Getreuer Yale University 2012

3) Tom.Goldstein “The Split Bregman Method For L1 Regularized Problems. "May 22 2008

4) Jos´e M. Bioucas-Dias M´ario A. T. Figueiredo Instituto de Telecomunicac¸ ˜oes,Instituto Superior T´ecnico,Lisboa, portugal “total variation restoration of speckled images using a split-bregman algorithm” march 24 2009

5) Rong-Qing Jia ∗, Hanqing Zhao, Wei Zhao “Convergence analysis of the Bregman method for the variational model of image denoising”

6) C.Y. Lee”Unconstrained Optimization”

7) “Basics of Unconstrained Optimization**”**-http://www.pages.drexel.edu/

8) Xiyin Li “Fundamentals of Constrained optimization”

9) Richard de Neufville, Joel Clark, and Frank R. Field Massachusetts Institute of Technology”constrained optimization**”**

10) Rong-Qing Jia, Hanqing Zhao, Wei Zhao”Convergence analysis of the Bregman method for the variational model ofimage denoising

11) ”Bregman Algorithms Author “JacquelineBush Supervisor Dr.Carlos Garc´ıa-Cervera” June 10, 2011

12) Rudin Osher Fatemi “Total Variation Denoising using Split Bregman’’Pascal Getreuer Yale University 2012

13) Tom.Goldstein “The Split Bregman Method For L1 Regularized Problems. "May 22 2008