Structured Total Least Squares and Image Processing

Jesse L. Barlow
The Pennsylvania State University
University Park, PA USA
barlow@cse.psu.edu

Joint work with

Haoying Fu (Ph.D. student, PSU)
Michael Ng (U. Hong Kong)

URL: `www.cse.psu.edu/~barlow/colortls4.ps`

Outline of Talk

- Least Squares and Total Least Squares
- Regularizing Least Squares and Total Least Squares
- Structured Total Least Squares
- An Errors–In–Variables Problem in Color Imaging

Least Squares from Gauss–Markov Model

In the deblurring problem in image processing, we suppose a linear relation with additive noise. That is

$$\mathbf{g} = H\mathbf{f} + \mathbf{r}$$

$\mathbf{g}$	observed image
$\mathbf{f}$	exact image
$\mathbf{r}$	unmeasurable error
H	blur matrix

Using Gauss–Markov assumptions we get

$$\min_{\mathbf{f}\in \mathsf{R}^n} \|\mathbf{g} - H\mathbf{f}\|_2$$

where

$$\|\mathbf{r}\|_2 \stackrel{def}{=} \left(\sum_{i=1}^{n} r_i^2 \right)^{1/2}$$

Least Squares (2)

Simple to solve

Classical linear algebra problem

- Q–R factorization
- Cholesky factorization of normal equations
- Iterative methods for large scale problems

Caveat – don't trust the solution

- The may be uncertainties in H. Two examples from image processing include
  - Camera alignments in superresolution problem
  - Weighting parameters in color imaging

  Lead to errors–in–variables problem.
- H is very ill-conditioned. Regularization is necessary.
- Overemphasis of outliers. ℓ-1 minimization may be advisable. See Fu et al. (2004). Leads to quadratic programming problem.

Total Least Squares (1)

Errors in variables model – Sprent (1966), Gleser (1982)

$$\mathbf{g} = (H + E)\mathbf{f} + \mathbf{r}$$

where we want to minimize

$$\|(\; E \quad \mathbf{r} \;)\|_F$$

Note –

$$\|A\|_F \stackrel{def}{=} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2} .$$

Better formulation for use is

$$\min_{\tilde{H},\mathbf{f}} \|\mathbf{g} - \tilde{H}\mathbf{f}\|_2^2 + \|\tilde{H} - H\|_F^2$$

This problem also has a nice linear algebra solution.

Total Least Squares (2)

Let

$$C = \begin{pmatrix} H & \mathbf{g} \end{pmatrix}$$

have the *singular value decomposition*

$$C = U\Sigma V^T$$

$$\begin{aligned} U \in \mathsf{R}^{m\times(n+1)}, &\qquad U^T U = I_{n+1}, \\ V \in \mathsf{R}^{(n+1)\times(n+1)}, &\qquad V^T V = VV^T = I_{n+1} \\ \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{n+1}) & \\ \sigma_1 \geq \sigma_2 \geq \ldots \sigma_{n+1} \geq 0. & \end{aligned}$$

Total Least Squares (3)

Take

$$U = (\mathbf{u}_1, \ldots, \mathbf{u}_{n+1}), \quad V = (\mathbf{v}_1, \ldots, \mathbf{v}_{n+1})$$

IF $\mathbf{e}_{n+1}^T \mathbf{v}_{n+1} \neq 0$, the solution is

$$\begin{bmatrix} \mathbf{f}_{TLS} \\ -1 \end{bmatrix} = -\frac{1}{\mathbf{e}_{n+1}^T \mathbf{v}_{n+1}} \mathbf{v}_{n+1}$$

Alteratively, $\mathbf{f}_{TLS}$ solves

$$(H^T H - \sigma_{n+1}^2 I)\mathbf{f}_{TLS} = H^T \mathbf{b}.$$

ELSE let

$$V = \begin{pmatrix} \overset{k}{V_1} & \overset{n-k+1}{V_2} \end{pmatrix}$$

where k is the smallest integer such that

$$V_2^T \mathbf{e}_{n+1} \neq 0.$$

Then

$$\begin{bmatrix} \mathbf{f}_{TLS} \\ -1 \end{bmatrix} = -\frac{1}{\|V_2^T \mathbf{e}_{n+1}\|_2^2} V_2 V_2^T \mathbf{e}_{n+1}.$$

Ill–Conditioning and Regularization

Many ways to handle it – truncated singular value decompostion and Krylov space methods. Simplest is probably Tikhanov regularization. Here we solve

$$\min_{\mathbf{f}} \|\mathbf{g} - H\mathbf{f}\|_2$$

subject to

$$\|L\mathbf{f}\|_2 \leq \alpha.$$

for some L and $\mathbf{f}$. Equivalent to assuming a prior distribution on $\mathbf{f}$.

For some $\lambda \geq 0$, this is equivalent to

$$\min_{\mathbf{f}\in R^n} \|\mathbf{g} - H\mathbf{f}\|_2^2 + \lambda\|L\mathbf{f}\|_2^2$$

This lambda may be found by solving a nonlinear equation.

Another common constraint is

$$\|\mathbf{g} - H\mathbf{f}\|_2 \leq \beta$$

L-curve criterion – tries to get modest values for both $\|\mathbf{g} - H\mathbf{f}\|_2$ and $\|L\mathbf{f}\|_2$.

Regularization of Total Least Squares

$$\min_{\mathbf{f},\tilde{H}} \|\mathbf{g} - H\mathbf{f}\|_2^2 + \|\tilde{H} - H\|_F^2$$

subject to

$$\|L\mathbf{f}\|_2 \leq \alpha.$$

Golub, Hansen, and O'Leary(1999)

For a two parameters $\lambda_L \geq 0$ and $\lambda_I \leq 0$, $\mathbf{f}$ is the solution of

$$(H^T H + \lambda_L L^T L + \lambda_I I)\mathbf{f} = H^T \mathbf{b}$$

Can be solved as a nonlinear equation in the two variables λ_I and λ_L.

See

- Guo and Renaut (1999,2004)
- Sima, Van Huffel, and Golub(2004)
- Fu and B. (2004)

Structured Total Least Squares (1)

Suppose that H is a structured matrix. Moreover, that structure is a big part of what the problem is about.

E.g. Toeplitz matrices – common in image restoration problems.

$$
\begin{aligned}
H &= \begin{bmatrix} t_0 & t_1 & t_2 & t_3 \\ t_{-1} & t_0 & t_1 & t_2 \\ t_{-2} & t_{-1} & t_0 & t_1 \\ t_{-3} & t_{-2} & t_{-1} & t_0 \end{bmatrix} \\
&= Toep(\mathbf{t})
\end{aligned}
$$

where

$$\mathbf{t} = (t_0, t_1, t_2, t_3, t_{-1}, t_{-2}, t_{-3})^T.$$

Choose $\tilde{H} = Toep(\tilde{\mathbf{t}})$.

Do

$$\|\tilde{H} - H\|_F = \|D(\tilde{\mathbf{t}} - \mathbf{t})\|_2$$

where D is a diagonal matrix.

Structured Total Least Squares (2)

With regularization (necessary under any common sense conditions) this becomes the nonlinear least squares problem

$$\min_{\mathbf{f},\tilde{\mathbf{t}}} \|\mathbf{g} - H(\tilde{\mathbf{t}})\mathbf{f}\|_2^2 + \|D(\mathbf{t} - \tilde{\mathbf{t}})\|_2^2 + \lambda\|L\mathbf{f}\|_2^2$$

General computational procedure – Rosen, Park, and Glick (1996)

Application to superresolution deblurring – Fu and B. (2004)

Color Imaging

Color Image $\mathbf{f}$

Three components, Red, Green, and Blue.

Let

$$f_r(x, y), \quad f_g(x, y), \quad f_b(x, y)$$

be the red, green, and blue components of the "true" scene.

Let

$$u_r(x, y), \quad u_g(x, y), \quad u_b(x, y)$$

be red, green , and blue components of the observed scene.

Define

$$h_{ij}(x, y), \quad i, j \in \{r, g, b\}$$

are the within channel and cross channel point spread functions (PSF) normalized to integrated to one.

Continuous and Discrete Models

$$\begin{aligned} u_r &= w_{rr}h_{rr} \star f_r + w_{gr}h_{gr} \star f_g + w_{br}h_{br} \star f_b + n_r \\ u_g &= w_{rg}h_{rg} \star f_r + w_{gg}h_{gg} \star f_g + w_{bg}h_{bg} \star f_g + n_g \\ u_b &= w_{rb}h_{rb} \star f_r + w_{gb}h_{gb} \star f_g + w_{bb}h_{bb} \star f_b + n_b. \end{aligned}$$

Here $\star$ donotes convolution. Thus the above is a Fredholm integral equation of the first kind. The $w_{ij}, i, j \in \{r, g, b\}$ are positive weights. Solving for them is part of the problem.

$\Downarrow$ Discretize!

$$\begin{aligned} \mathbf{u}_r &= w_{rr}H_{rr}\mathbf{f}_r + w_{gr}H_{gr}\mathbf{f}_g + w_{br}H_{br}\mathbf{f}_b + \mathbf{n}_r \\ \mathbf{u}_g &= w_{rg}H_{rg}\mathbf{f}_r + w_{gg}H_{gg}\mathbf{f}_g + w_{bg}H_{bg}\mathbf{f}_g + \mathbf{n}_g \\ \mathbf{u}_b &= w_{rb}H_{rb}\mathbf{f}_r + w_{gb}H_{gb}\mathbf{f}_g + w_{bb}H_{bb}\mathbf{f}_b + \mathbf{n}_b. \end{aligned}$$

Here H_{ij} the within channel or cross channel blur matrix and $\mathbf{n}_r$, $\mathbf{n}_g$ and $\mathbf{n}_b$ are additive noise.

Matrix Version of the Model

In good matrix form

$$\mathbf{u} = H(\mathbf{w})\mathbf{f} + \mathbf{n}$$

$$H(\mathbf{w}) = H(\mathbf{w}) = \begin{bmatrix} w_{rr}H_{rr} & w_{gr}H_{gr} & w_{br}H_{br} \\ w_{rg}H_{rg} & w_{gg}H_{gg} & w_{bg}H_{bg} \\ w_{rb}H_{rb} & w_{gb}H_{gb} & w_{bb}H_{bb} \end{bmatrix}.$$

$$\mathbf{w} = (w_{rr}, w_{gr}, w_{br}, w_{rg}, w_{gg}, w_{bg}, w_{rb}, w_{gb}, w_{bb})^T,$$

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_r \\ \mathbf{f}_g \\ \mathbf{f}_b \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} \mathbf{u}_r \\ \mathbf{u}_g \\ \mathbf{u}_b \end{bmatrix}, \quad \text{and } \mathbf{n} = \begin{bmatrix} \mathbf{n}_r \\ \mathbf{n}_g \\ \mathbf{n}_b \end{bmatrix},$$

Boundary Conditions

Zero Boundary Conditions

H_{ij} is Block Toeplitz with Toeplitz Blocks.

Not a terribly good boundary condition. It tends to be inaccurate at the edges.

Neumann Boundary Conditions

H_{ij} is Block Toeplitz plus Hankel with Toeplitz plus Hankel Blocks (BTHTHB)

These are pretty useful BC.

Toeplitz and Hankel Matrices – In 4×4 Examples

Toeplitz

$$T = \begin{bmatrix} t_0 & t_1 & t_3 & t_4 \\ t_{-1} & t_0 & t_1 & t_2 \\ t_{-2} & t_{-1} & t_0 & t_1 \\ t_{-3} & t_{-2} & t_{-1} & t_0 \end{bmatrix}$$

It is banded in many applications.

Block Toeplitz

$$T = \begin{bmatrix} T_0 & T_1 & T_3 & T_4 \\ T_{-1} & T_0 & T_1 & T_2 \\ T_{-2} & T_{-1} & T_0 & T_1 \\ T_{-3} & T_{-2} & T_{-1} & T_0 \end{bmatrix}$$

where T_j is a matrix of appropriate dimension. It is BTTB is the blocks are Toeplitz.

Hankel

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ h_2 & h_3 & h_4 & h_1 \\ h_3 & h_4 & h_1 & h_2 \\ h_4 & h_1 & h_2 & h_3 \end{bmatrix}$$

You get the idea what BTHTHB would mean.

Matrix–vector muliplication with these matrices is fast, about $O(n \log n)$.

Ill–Conditioning

Fredholm integral equations of the first kind are ill–posed, in general.

For a fixed $\mathbf{w}$

$$\min_{\mathbf{f}} \|\mathbf{u} - H(\mathbf{w})\mathbf{f}\|_2^2$$

does not obtain a meaningful solution. The ill-posedness of the original equation makes $H(\mathbf{w})$ very ill–conditioned.

Thus we need to regularize the solution. Add penalty terms.

Regularization

Galatsanos et al (1991) proposed the following regularization scheme. We add the extra variable $\mathbf{z}$.

$$\min_{\mathbf{f},\mathbf{z}} \|\mathbf{u} - H(\mathbf{w}+\mathbf{z})\mathbf{f}\|_2^2 + \|\sqrt{\Omega}R(\mathbf{f})\mathbf{f}\|_2^2 + \alpha_h\|\mathbf{z}\|_2^2$$

$$R(\mathbf{f}) = DLD^{-1},$$

L is a 3-D Laplacian.

$D = \text{diag}(\|\mathbf{f}_r\|_2 I, \|\mathbf{f}_g\|_2 I, \|\mathbf{f}_p\|_2 I), \quad \Omega = \text{diag}(\alpha_r I, \alpha_g I, \alpha_p I)$

where α_g,α_g and α_b are regularization parameters.

The last term is an acknowledgement that $\mathbf{w}$ and $\mathbf{w}+\mathbf{z}$ must be "close" together. In other words, $\mathbf{w}$ is a reasonable estimate of the true weight vector.

Add a Constraint

Also, $\mathbf{z}$ is constrained by

$$\sum_i (\mathbf{w}+\mathbf{z})_{ij} = \sum_i \mathbf{w}_{ij}$$

That allows us to eliminate three components, and let

$$\hat{\mathbf{z}} = [z_{rr} \quad z_{gr} \quad z_{rg} \quad z_{gg} \quad z_{rb} \quad z_{gb}]^T,$$

Thus

$$\mathbf{z} = S\hat{\mathbf{z}}$$

Non–Linear Least Squares Procedure

Thus the problem we want to solve is

$$\min_{\mathbf{f},\widehat{\mathbf{z}}} \|H(\mathbf{w}+\mathbf{z})\mathbf{f}-\mathbf{u}\|_2^2 + \|\sqrt{\Omega}R(\mathbf{f})\mathbf{f}\|_2^2 + \alpha_h\|S\widehat{\mathbf{z}}\|_2^2.$$

This is a nonlinear least squares problem. We solve it by a Gauss–Newton iteration.

Rosen, Park, and Glick (1996) – Structured Total Least Squares (STLS)

Preussner and O'Leary (2003) – Applied STLS to Image Deblurring

Fu and Barlow (2004) – Applied STLS to Superresolution Problem

Gauss–Newton Step for STLS Procedure

A Gauss–Newton step for this STLS procedure is

$$\min_{\Delta \mathbf{f}, \Delta \widehat{\mathbf{z}}} \left\| \begin{bmatrix} H(\mathbf{w}+\mathbf{z}) & Z \\ \sqrt{\Omega} R(\mathbf{f}) & 0 \\ 0 & \sqrt{\alpha_h} S \end{bmatrix} \begin{bmatrix} \Delta \mathbf{f} \\ \Delta \widehat{\mathbf{z}} \end{bmatrix} - \begin{bmatrix} \mathbf{u} - H(\mathbf{w}+\mathbf{z})\mathbf{f} \\ -\sqrt{\Omega} R(\mathbf{f})\mathbf{f} \\ -\sqrt{\alpha_h} S \widehat{\mathbf{z}} \end{bmatrix} \right\|_2 .$$

where Z is a block diagonal matrix such that

$$H(\mathbf{w}+\mathbf{z})\Delta \mathbf{f} = Z \Delta \mathbf{z}.$$

This problem is LARGE. The Z and S matrices are sparse and H is structured. It must be solved by Krylov space methods, e.g. LSQR or CGLS. That is a conjugate gradient for least squares.

Preconditioning

Thus we have to find a preconditioner for solving

$$\min \|\mathbf{b} - A\mathbf{x}\|_2$$

where

$$A = \begin{bmatrix} H(\mathbf{w}+\mathbf{z}) & Z \\ \sqrt{\Omega}R(\mathbf{f}) & 0 \\ 0 & \sqrt{\alpha_h}S \end{bmatrix}$$

and Z and S are simple matrices.

We look at the normal equations matrix

$$A^T A = \begin{bmatrix} \widehat{M} & H(\mathbf{w}+\mathbf{z})^T Z \\ Z^T H(\mathbf{w}+\mathbf{z}) & Z^T Z + \alpha_h S^T S \end{bmatrix}.$$

where

$$\widehat{M} = H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) + R(\mathbf{f})^T \Omega R(\mathbf{f})$$

and make use of the fact that much is known about how to precondition $H(\mathbf{w}+\mathbf{z})$.

Preconditioning a Symmetric PSF (1)

A PSF is symmetric if

$$h(x,y) = h(x,-y) = h(-x,y) = h(-x,-y)$$

In that case, from a result of Ng et al. (1999) H can be reduced using the discrete cosine transform (DCT).

$$H = (I \otimes C^T) \begin{bmatrix} \Lambda_{rr} & \Lambda_{gr} & \Lambda_{br} \\ \Lambda_{rg} & \Lambda_{gg} & \Lambda_{bg} \\ \Lambda_{rb} & \Lambda_{gb} & \Lambda_{bb} \end{bmatrix} (I \otimes C)$$

Likewise,

$$R(\mathrm{f}) = D^{-1}\widehat{C}^T \begin{bmatrix} \Lambda_L + 2I & -I & -I \\ -I & \Lambda_L + 2I & -I \\ -I & -I & \Lambda_L + 2I \end{bmatrix} \widehat{C} D$$

where $\widehat{C} = I \otimes C$.

Preconditioning a Symmetric PSF (2)

Thus

$$\widehat{M} = \widehat{C}^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{12} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{13} & \Lambda_{23} & \Lambda_{33} \end{bmatrix} \widehat{C}$$

Apply a permutation and we have that

$$\widehat{M} = \tilde{C}^T \text{diag}(B_{i,j}) \tilde{C}$$

where $\tilde{C} = P\widehat{C}$ and $i, j \in \{1, 2, \ldots, n\}$. Under weak and reasonable assumptions, this matrix is positive definite.

The inverse is

$$\widehat{M}^{-1} = \tilde{C}^T \text{diag}(B_{i,j}^{-1}) \tilde{C}.$$

Moreover, $M = G^T G$ where

$$G = \text{diag}(G_{i,j}) \tilde{C}.$$

Can be obtained without producing $\widehat{M}$ using Q–R factorization.

Preconditioning a Symmetric PSF (3)

Choose the preconditioner

$$M = \begin{bmatrix} \widehat{M} & 0 \\ 0 & Z^T Z + \alpha_h S^T S \end{bmatrix}.$$

Then

$$M^{-1} A^T A = I + E_1$$

where E_1 has rank 12. Guaranteed convergence in 13 iterations.

Preconditioning a General PSF (1)

Substitute the PSF

$$\mathbf{h}_s(i,j) = (\mathbf{h}(i,j) + \mathbf{h}(i,-j) + \mathbf{h}(-i,j) + \mathbf{h}(-i,-j))/4$$

with the Neumann BC. Define

$$\delta = \sum_{i,j} |h(i,j) - h_s(i,j)|$$

Create new blurring matrix $c(H)$ based upon h_s. Then

$$\|H - c(H)\| \leq 4\delta$$

and the 1,2 and ∞ norms.

Let

$$\begin{aligned} M_c &= \begin{bmatrix} \widehat{M}_c & 0 \\ 0 & Z^T Z + \alpha_h S^T S \end{bmatrix} \\ \widehat{M}_c &= c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z})) + R(\mathbf{f})^T \Omega R(\mathbf{f}) \end{aligned}$$

Preconditioning a General PSF (2)

We can show that

$$M_c^{-1} A^T A = I + E_1 + E_2$$

where E_1 has rank 12 and

$$\|E_2\|_2 \leq 8\delta \|\mathbf{w} + \mathbf{z}\|_\infty \|M_c^{-1}\|_2.$$

Thus, again we expect rapid convergence.

A Computational Example (1)

We used the following PSFs.

$$\mathbf{h}_{rr} = \mathbf{h}_{gr} = \mathbf{h}_{br} = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix},$$

$$\mathbf{h}_{rg} = \mathbf{h}_{gg} = \mathbf{h}_{bg} = \begin{bmatrix} 0 & 0.125 & 0 \\ 0.125 & 0.5 & 0.125 \\ 0 & 0.125 & 0 \end{bmatrix},$$

$$\mathbf{h}_{rb} = \mathbf{h}_{gb} = \mathbf{h}_{bb} = \begin{bmatrix} 0.0625 & 0.125 & 0.0625 \\ 0.125 & 0.25 & 0.125 \\ 0.0625 & 0.125 & 0.0625 \end{bmatrix}.$$

We added random noise of size about 0.01 is added to these PSFs and then they are normalized to integrate to one. This makes them slightly nonsymmetric.

For both least squares (fixed $\mathbf{w}$) and total least squares solutions we used $[\alpha_r \quad \alpha_g \quad \alpha_b] = [0.0001 \quad 0.0001 \quad 0.0001]$.

Example – The Weights

The weights used to generate the observed channels are as follows:

$$\mathbf{w} = [0.5, 0.3, 0.2, 0.25, 0.5, 0.25, 0.2, 0.3, 0.5]^T.$$

A noise returned by the MATLAB command "0.01*(rand-0.5)" is added to the observed channels. The estimated weights are set to

$$\begin{aligned}\mathbf{w} = \; & [0.45, 0.33, 0.22, 0.275, 0.45, 0.275, \\ & 0.22, 0.33, 0.55]^T.\end{aligned}$$

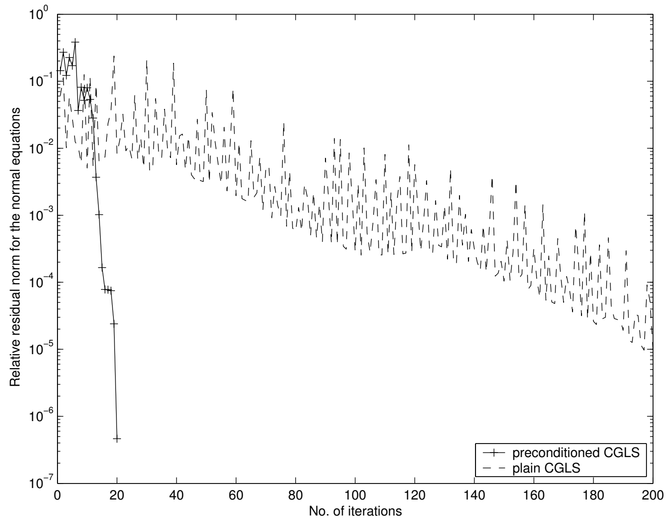
The relative error on the estimated weights is 10%.

For the TLS solution we used $\alpha_h = 9$. The recovered weights were

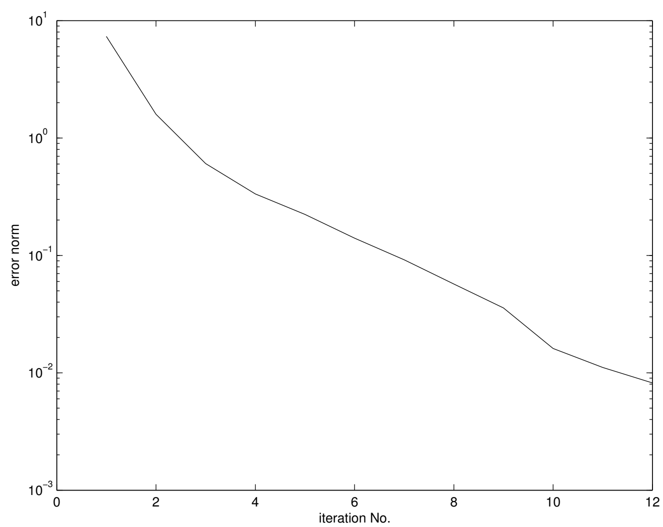
$$\begin{aligned}\mathbf{w} = [& 0.4943, 0.3137, 0.1920, 0.2555, 0.4688, \\ & 0.2757, 0.1902, 0.3292, 0.4807]^T.\end{aligned}$$

a considerable improvement.

Example – Convergence



Conjugate Gradient convergence



TLS algorithm convergence

Example – Original and Observed Images

The original image

The observed image

Example – Least Squares Solutions

The image recovered by the least squares method. The observed image is used to scale $R(\mathrm{f})$. PSNR=75.92

The image recovered by the least squares method. The original image is used to scale $R(\mathrm{f})$. PSNR=76.14

Example – TLS Solutions

The image recovered by the TLS Algorithm, PSNR=82.35

The image recovered by a variant of the TLS Algorithm (a fixed $R(\mathrm{f})$ is used), PSNR=81.53

Conclusion

- The image deblurring problem requires consideration of a number of alternatives to the Gauss–Markov linear model.

- The errors–in–variables model for computing the inter channel and cross channel weights leads to the recovery of a better color image.

- Problem is easily formulated and solved as a nonlinear least squares problem.

- A Gauss–Newton step for the nonnlinear least squares problem can be performed using preconditioned conjugate gradient. The precoditioner of choice is one that is based upon the discrete cosine transform.