Some work notes on adding a bias term on a BD problem

Zhengyu Chen

June 25, 2015

Abstract

This note is base on the work note of Blind deconvolution problem written by Yenson. To apply this model on the STS system, we need to add a bias term onto our optimization problem, which is caused in the measurement. As a result of that, we need to do some reformulating on our object function so that we could use the similar method to solve the problem.

1 An optimization problem for STS appliction

1.1 Notation

We use most of the notations from Yenson's note and add a new term of $\beta \in \mathbb{R}^N$. The index N is over $[N] = 0, 1, ..., N-1$ and the vector is

$$\beta = [\beta_0, \beta_1, ..., \beta_{N-1}]^\top$$

1.2 Problem setting

Same as the former problem, we need to investigate the problem

$$\min_{\|A\|^2=1} \varphi^\lambda(A), \tag{1}$$

where

$$\begin{aligned} \varphi^\lambda(A) &= \min_{X,\beta} \psi^\lambda(A, X, \beta) &(2)\\ \psi^\lambda(A, X, \beta) &= \frac{1}{2}\sum_{i=0}^{N-1} \|Y_i - A_i * X - \beta_i\|_2^2 + \lambda r(X) &(3) \end{aligned}$$

similarly as former notes, with some sparse penalty $r(X)$, the solution of Problem (1) will be given by

$$\begin{aligned} \hat{A}_i &= \arg\min_{\|A_i\|=1} \varphi^\lambda(A_i), &(4)\\ [X, \beta]^*(A) &= \arg\min_{[X,\beta]} \psi^\lambda(A, X, \beta) &(5) \end{aligned}$$

2 Frist and second order properties of object

Since the they both stand in a euclid space, we could treat X and β into one vector $W \in \mathbb{R}^{m+N}$ where

$$W = \begin{bmatrix} X \\ \beta \end{bmatrix} \tag{6}$$

2.1 Jacobians

Due to the calculaor of vector, we could got the Jacobian of W as

$$\frac{\mathrm{d}f}{\mathrm{d}W} = \begin{bmatrix} \frac{\partial f}{\partial X_0} \\ \frac{\partial f}{\partial X_1} \\ \vdots \\ \frac{\partial f}{\partial X_{m-1}} \\ \frac{\partial f}{\partial \beta_0} \\ \vdots \\ \frac{\partial f}{\partial \beta_{N-1}} \end{bmatrix} \tag{7}$$

and Jacobian of A as

$$\nabla_{\mathrm{A}}\varphi = \begin{bmatrix} \nabla_{\mathrm{A_0}}\varphi \\ \vdots \\ \nabla_{\mathrm{A_{N-1}}}\varphi \end{bmatrix} \tag{8}$$

2.2 Gradients and Hessians of ψ

2.2.1 Dealing with W

Firstly, we reformulate the object function (3) as

$$\begin{aligned} \psi(A, X, \beta) &= \frac{1}{2}\sum_{i=0}^{N-1} \|Y - A_i - \beta \mathbb{1}_i\|_2^2 + \lambda r(X) &(9)\\ &= \frac{1}{2}\sum_{i=0}^{N-1} \left\| \begin{bmatrix} \mathbf{C}_{\widetilde{A_i}} & \mathbb{1}_i \end{bmatrix} \begin{bmatrix} X \\ \beta \end{bmatrix} - Y_i \right\|_2^2 + \lambda r(X) &(10) \end{aligned}$$

where

$$\mathbb{1}_i = \begin{bmatrix} \underbrace{\mathbf{0} \ldots \mathbf{0}}_{i\ colums} & \mathbf{1} & \ldots & \mathbf{0} \end{bmatrix} \tag{11}$$

Gradient of W is

$$\nabla_w \psi = \sum_{i=0}^{N-1} \begin{pmatrix} \mathbf{C}^*_{\widetilde{A}_i} \\ \mathbb{1}^*_i \end{pmatrix} \left(\begin{bmatrix} \mathbf{C}_{\widetilde{A}_i} & \mathbb{1}_i \end{bmatrix} \begin{pmatrix} X \\ \beta \end{pmatrix} - Y_i \right) + \lambda r(X) \cdot \begin{pmatrix} \mathbf{I} \\ \mathbf{0} \end{pmatrix} \tag{12}$$

$$= \sum_{i=0}^{N-1} \begin{pmatrix} \overleftarrow{A}_i \circledast g_i \\ 0 \\ \vdots \\ \sum g_i \\ \vdots \\ 0 \end{pmatrix} + \lambda r(X) \cdot \begin{pmatrix} \mathbf{I} \\ \mathbf{0} \end{pmatrix} \tag{13}$$

where

$$g_i = \left(\begin{bmatrix} \mathbf{C}_{\widetilde{A}_i} & \mathbb{1}_i \end{bmatrix} \begin{pmatrix} X \\ \beta \end{pmatrix} - Y_i \right) \tag{14}$$

The Hessian is

$$\nabla^2_{ww} \psi = \sum_{i=0}^{N-1} \begin{pmatrix} \mathbf{C}^*_{\widetilde{A}_i} \\ \mathbb{1}^*_i \end{pmatrix} \begin{pmatrix} \mathbf{C}_{\widetilde{A}_i} & \mathbb{1}_i \end{pmatrix} + \begin{pmatrix} \lambda \nabla^2_{xx} \psi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{15}$$

2.2.2 Dealing with A_i

To deal with A_i, we need to reformulate the object function as:

$$\psi = \sum_{i=0}^{N-1} \|\mathbf{C}_X \mathbf{I}_{k\to m} A_i + \mathbb{1}_i \beta - Y_i\|_2^2 + \lambda r(X) \tag{16}$$

The gradient of A_i is

$$\nabla_{A_i} = \mathbf{I}^*_{k\to m} \mathbf{C}^*_X (\mathbf{C}_X \mathbf{I}_{k\to m} A_i + \mathbb{1}_i \beta - Y_i) \tag{17}$$

and the hessian of A_i is

$$\nabla^2_{A_i A_j} = \begin{cases} \mathbf{I}^*_{k\to m} \mathbf{C}^*_X \mathbf{C}_X \mathbf{I}_{k\to m} & \text{if } i = j \\ \mathbf{0} & \text{if } i \neq j \end{cases} \tag{18}$$

2.2.3 Dealing with cross term

From the formulation (17) we could get that

$$\nabla_{A_i} = \mathbf{I}^*_{k\to m} \overleftarrow{X} \circledast (X \circledast \widetilde{A}_i + \mathbb{1}_i \beta - Y_i) \tag{19}$$

So the cross hessian $\nabla^2_{A_i W}$ could be written like

$$\nabla^2_{A_i W} = \mathbf{I}^*_{k\to m} \frac{\mathrm{d}}{\mathrm{d}W} \overleftarrow{X} \circledast (X \circledast \widetilde{A}_i + \mathbb{1}_i \beta - Y_i) \tag{20}$$

$$= \mathbf{I}^*_{k\to m} (\underbrace{\frac{\mathrm{d}}{\mathrm{d}W} \overleftarrow{X} \circledast (X \circledast \widetilde{A}_i + \mathbb{1}_i \beta)}_{Part\ 1} - \underbrace{\frac{\mathrm{d}}{\mathrm{d}W} \mathbf{C}_{Y_i} \overleftarrow{\Pi} X}_{Part\ 2}) \tag{21}$$

$Part1$ could be written as

$$\frac{\mathrm{d}}{\mathrm{d}W}(\underbrace{\overleftarrow{X} \circledast \mathbf{C}_{\widetilde{A}_i}}_{Part1.1} + \underbrace{\overleftarrow{X} \circledast \mathbb{1}_i \beta}_{Part1.2}) \tag{22}$$

Where $Part1.1$ could be easily get

$$\begin{bmatrix} \mathbf{C}_{\widetilde{A}_i}(\mathbf{C}^*_X + \mathbf{C}_X \overleftarrow{\Pi}) & \mathbf{0} \end{bmatrix} \tag{23}$$

and $Part1.2$ is

$$\frac{\mathrm{d}}{\mathrm{d}W}\overleftarrow{X} \circledast \mathbb{1}_i \beta \tag{24}$$

$$= \frac{\mathrm{d}}{\mathrm{d}W}\mathbf{C}^*_X \mathbb{1}_i \beta + \frac{\mathrm{d}}{\mathrm{d}W}\mathbf{C}_{\mathbb{1}_i \beta}\overleftarrow{\Pi} X \tag{25}$$

$$= [\mathbf{0} \quad \mathbf{C}^*_X \mathbb{1}_i] + \begin{bmatrix} \mathbf{C}_{\mathbb{1}_i \beta}\overleftarrow{\Pi} & \mathbf{0} \end{bmatrix} \tag{26}$$

Put all of above together, we could find the result:

$$\nabla^2_{A_i W}\psi = \mathbf{I}^*_{k \to m}\left(\begin{bmatrix} \mathbf{C}_{\widetilde{A}_i \circledast \overleftarrow{X}} + \mathbf{C}_{\widetilde{A}_i \circledast X + \mathbb{1}_i \beta - Y_i}\overleftarrow{\Pi} & \mathbf{C}^*_X \mathbb{1}_i \end{bmatrix}\right) \tag{27}$$

It could be easily verify that

$$\nabla^2_{W A_i}\psi = (\nabla^2_{A_i W})^* \psi \tag{28}$$

2.3 Gradient and Hassian of φ

To the first order gradient, we have

$$\nabla_{A_i}\varphi = \nabla_{A_i}\psi \tag{29}$$

As a result we could got,

$$\nabla_A \varphi = \nabla_A \min_{X,\beta} \psi(A, X, \beta) \tag{30}$$

$$= \begin{bmatrix} \nabla_{A_0}\varphi \\ \vdots \\ \nabla_{A_{N-1}}\varphi \end{bmatrix} \tag{31}$$

The hessian of φ is

$$\nabla^2_{AA}\varphi = \nabla^2_{AA}\psi - \nabla^2_{AW}\psi(\nabla^2_{WW}\psi)^{-1}\psi\nabla^2_{WA}\psi \tag{32}$$

where

$$\nabla^2_{AA}\psi = \nabla_A \nabla_A \psi \tag{33}$$

$$= \begin{pmatrix} \nabla^2_{A_0 A_0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \nabla^2_{A_1 A_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \nabla^2_{A_{N-1} A_{N-1}} \end{pmatrix} \tag{34}$$