

# Block Relaxation as Majorization

Jan de Leeuw

Version 007, November 04, 2016

## Abstract

This short note shows that all block relaxation algorithms can be formulated as majorization algorithms. The result is mostly a curiosity, without any obvious practical applications.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Block Relaxation</b>	<b>2</b>
<b>3</b>	<b>Augmentation</b>	<b>2</b>
<b>4</b>	<b>Majorization</b>	<b>2</b>
<b>5</b>	<b>Majorization from Blocking</b>	<b>3</b>
<b>6</b>	<b>Partial Majorization</b>	<b>4</b>
	<b>References</b>	<b>4</b>

---

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory `deleeuwpx/pubfolders/block` has a pdf version, the complete Rmd file, and the bib file.

## 1 Introduction

We use notation and terminology taken from De Leeuw (1994).

## 2 Block Relaxation

To minimize  $g : X \otimes Y \rightarrow \mathbb{R}$  over  $x \in X$  and  $y \in Y$  we can use the *block relaxation* algorithm.

$$\begin{aligned} y^{(k+1)} &\in \mathbf{argmin}_{y \in Y} g(x^{(k)}, y), \\ x^{(k+1)} &\in \mathbf{argmin}_{x \in X} g(x, y^{(k+1)}). \end{aligned}$$

Note that the argmin's are point-to-set maps, because the minima over blocks are not necessarily unique.

As an example, consider  $g(a, b) = \mathbf{SSQ}(y - Xa - Zb)$  with  $\mathbf{SSQ}()$  the sum of squares. The algorithm, using Moore-Penrose inverses, is

$$\begin{aligned} b^{(k+1)} &= Z^+(y - Xa^{(k)}), \\ a^{(k+1)} &= X^+(y - Zb^{(k+1)}). \end{aligned}$$

## 3 Augmentation

Suppose the original problem is to minimize  $f : X \rightarrow \mathbb{R}$  over  $x \in X$  and we can find  $g : X \times Y \rightarrow \mathbb{R}$  such that  $f(x) = \min_{y \in Y} g(x, y)$ . Such a  $g$  is called an *augmentation* of  $f$ . Minimizing  $f$  over  $x \in X$  can be done by applying block relaxation to the augmentation  $g$  over  $x \in X$  and  $y \in Y$ .

In least squares factor analysis, for example, we minimize

$$f(X) = \mathbf{SSQ}(\mathbf{off}(R - XX')),$$

where  $\mathbf{off}(X) = X - \mathbf{diag}(X)$ . Choose the augmentation

$$g(X, \Delta) = \mathbf{SSQ}(R - XX' - \Delta)$$

where  $\Delta$  varies over diagonal matrices. The block relaxation algorithm is

$$\begin{aligned} \Delta^{(k+1)} &= \mathbf{diag}(R - X^{(k)}(X^{(k)})'), \\ (R - \Delta^{(k+1)})X^{(k+1)} &= X^{(k+1)}\Lambda, \end{aligned}$$

where  $\Lambda$  is a symmetric matrix of Lagrange multipliers. Thus finding  $X^{(k+1)}$  involves solving the eigen problem for  $R - \Delta^{(k+1)}$ .

## 4 Majorization

Again we want to minimize  $f : X \rightarrow \mathbb{R}$  over  $x \in X$ . Suppose there is a  $g : X \times X \rightarrow \mathbb{R}$  such that  $g(x, y) \geq f(x)$  for all  $x \in X$  and  $y \in X$  and such that  $g(x, x) = f(x)$  for all  $x \in X$ .

Such a  $g$  is called a *majorization* of  $f$ . Minimize  $f$  over  $x \in X$  by applying block relaxation to the majorization  $g$  over  $x \in X$  and  $y \in X$ .

Clearly any majorization of  $f$  is also an augmentation of  $f$ . Majorization is a special type of augmentation because  $X = Y$  and  $x \in \mathbf{argmin}_{y \in Y} g(x, y)$ . Thus the block relaxation is simply

$$x^{(k+1)} \in \mathbf{argmin}_{x \in X} g(x, x^{(k)}).$$

Thus majorization algorithms are block relaxation algorithms.

## 5 Majorization from Blocking

Suppose  $h : X \otimes Z \rightarrow \mathbb{R}$ . Define  $T(x) = \mathbf{argmin}_{z \in Z} h(x, z)$ , and suppose  $t(x)$  is a selection from  $T(x)$ , i.e.  $t(x) \in T(x)$  for all  $x \in X$ . Define  $f(x) = h(x, t(x))$  and  $g(x, y) = h(x, t(y))$ . Then  $g(x, y) \geq g(x, x) = f(x)$ . Thus  $g$  is a majorization of  $f$ . The majorization algorithm for  $f$  and  $g$  is simply the block relaxation algorithm for  $h$ . Thus block relaxation algorithms are majorization algorithms. Our reasoning here is very similar to Lange (2016 (in press)) (section 4.9).

As an example consider

$$h(X, \Delta) = \mathbf{SSQ}(R - XX' - \Delta).$$

Then

$$f(X) = \mathbf{SSQ}(\mathbf{off}(R - XX')),$$

and the majorization of  $f$  is

$$g(X, Y) = \mathbf{SSQ}(R - XX' - \mathbf{diag}(R - YY')).$$

Another example is

$$h(a, b) = \mathbf{SSQ}(y - Xa - Zb).$$

Then

$$f(a) = (y - Xa)'(I - ZZ^+)(y - Xa),$$

and the majorization of  $f$  is

$$g(a, b) = \mathbf{SSQ}(y - Xa - ZZ^+(y - Xb)).$$

Clearly we can also interchange the role of the two blocks. In the factor analysis example we can minimize out  $X$  to get

$$f(\Delta) = \sum_{s=p+1}^n \lambda_s(R - \Delta),$$

where the  $\lambda_s(X)$  are the ordered eigenvalues of  $X$  (assuming the  $p$  largest eigenvalues are non-negative). The majorization function is

$$g(\Delta, \Omega) = \mathbf{SSQ}(R - \Delta - (R - \Omega)_p),$$

with  $(X)_p$  the best rank  $p$  approximation of  $X$ .

## 6 Partial Majorization

Suppose the problem we want to solve is minimizing  $g(x, y)$  over  $x \in X$  and  $y \in Y$ . If both minimizing  $g(x, y)$  over  $x \in X$  for fixed  $y \in Y$  and minimizing  $g(x, y)$  over  $y \in Y$  for fixed  $x \in X$  is easy, then we often use block-relaxation, alternating the two conditional minimization problems until convergence.

But now suppose only one of the two problems, say minimizing  $g(x, y)$  over  $y \in Y$  for fixed  $x \in X$ , is easy. Define

$$f(x) = \min_{y \in Y} g(x, y)$$

and let  $y(x)$  be any  $y \in Y$  such that  $f(x) = g(x, y(x))$ .

Suppose we have a majorizing function  $h(x, z)$  for  $f(x)$ . Thus

$$\begin{aligned} f(x) &\leq h(x, z) && \forall x, z \in X, \\ f(x) &= h(x, x) && \forall x \in X. \end{aligned}$$

Suppose our current best solution for  $x$  is  $\tilde{x}$ , with corresponding  $\tilde{y} = y(\tilde{x})$ . Let  $x^+$  be any minimizer of  $h(x, \tilde{x})$  over  $x \in X$ . Now

$$g(x^+, y(x^+)) = f(x^+) \leq h(x^+, \tilde{x}) \leq h(\tilde{x}, \tilde{x}) = f(\tilde{x}) = g(\tilde{x}, y(\tilde{x}))$$

which means that  $(x^+, y(x^+))$  gives a lower loss function value than  $(\tilde{x}, y(\tilde{x}))$ . Thus we have, under the usual conditions, a convergent algorithm.

## References

- De Leeuw, J. 1994. "Block Relaxation Algorithms in Statistics." In *Information Systems and Data Analysis*, edited by H. H. Bock, W. Lenski, and M. M. Richter, 308–24. Berlin: Springer Verlag.
- Lange, K. 2016 (in press). *MM Optimization Algorithms*.