

Convergence Rate of Majorization Algorithms with Constraints

Jan de Leeuw

Version August 30, 2018

Abstract

dodo

Contents

1	Introduction	1
2	No Constraints	2
2.1	Example: SMACOF	3
3	Equality Constraints	4
3.1	Example: Power Method	5
3.2	Example: Least Squares with Quadratic Constraint	7
4	Parametric Equality Constraints	8
5	Inequality Constraints, Part I	9
6	Inequality Constraints, Part II	10
	References	10

Note: This is a working paper which will be expanded/updated frequently. All suggestions for improvement are welcome. The directory gifi.stat.ucla.edu/constraints has a pdf version, the bib file, the complete Rmd file, and the R and C code (if applicable).

1 Introduction

Majorization algorithms or **MM algorithms** are an increasingly popular class of techniques for iteratively solving the problem of minimizing a real valued function f over a set $\mathcal{X} \subseteq \mathbb{R}^n$. Some of the pertinent early references are De Leeuw (1994) and Heiser (1995), and there are recent book-length treatments in De Leeuw (2016) and Lange (2016).

A majorization algorithm is defined by a **majorization scheme**, which is a function $g : \mathcal{X} \otimes \mathcal{X} \Rightarrow \mathbb{R}$ such that $f(x) \leq g(x, y)$ for all $x, y \in \mathcal{X}$ and $g(x, x) = f(x)$ for all $x \in \mathcal{X}$ (De Leeuw (2016)).

$$\begin{aligned} \mathcal{X} : \mathbb{R}^n &\Rightarrow \mathbb{R}^n \\ f(x) &\leq g(x, y) \text{ for all } x \in \mathcal{X}(y) \end{aligned}$$

The **update map** of a majorization scheme is

$$\omega(y) \triangleq \mathbf{argmin} \{g(x, y) \mid x \in \mathcal{X}(y)\}. \quad (1)$$

Iteration $k + 1$ of the majorization algorithm updates the current best solution by the rule

$$x^{(k+1)} \in \omega(x^{(k)}).$$

Note that in general the update map is set-valued, but in the implementation of an actual algorithm we use a single valued selection from ω .

Convergence of majorization algorithms to a fixed point $y \in \omega(y)$ is guaranteed under conditions that are satisfied in many applications (De Leeuw (2016), Lange (2016)). For most majorization algorithms used so far convergence is linear, with a convergence rate equal to the spectral norm, i.e. the eigenvalue of maximum modulus, of the derivative $\mathcal{D}\omega(y)$ at y .

There is a huge literature on derivatives of the update map in **parametric mathematical programming**. Good overviews of the initial developments of the field are in Bank et al. (1982) and Fiacco (1983). Levitin (1994) is a truly virtuoso and exhaustive treatment, primarily using the tools of classical analysis. Recent modern book length treatments, using the more abstract framework of modern variational analysis, are in Bonnans and Shapiro (2000), Luderer, Minchenko, and Satsura (2002), and Dontchev and Rockafellar (2014). Excellent introductory overviews of the field are Giorgi and Zuccotti (1918) and the lecture notes of Still (2018).

2 No Constraints

If there are no constraints on x the derivative of the update map in a majorization algorithm follows from a straightforward application of the implicit function theorem (IFT). The result below, which applies the IFT to majorization algorithms, is taken from De Leeuw (2016).

Define $\mathcal{X} = \mathbb{R}^n$ and ω is define by (1). We assume the minimum exists and is unique, and that g is two times continuously differentiable in a neighborhood of $(\omega(y), y)$.

At a minimum

$$\mathcal{D}_1 g(\omega(y), y) = 0$$

Differentiating again

$$\mathcal{D}_{11} g(\omega, y) \mathcal{D}\omega(y) + \mathcal{D}_{12} g(\omega, y),$$

and thus, assuming the inverse exists,

$$\mathcal{D}\omega(y) = -[\mathcal{D}_{11}g(\omega(y), y)]^{-1}\mathcal{D}_{12}g(\omega(y), y).$$

So far, nothing new. This is just the usual formula for the derivatives of an implicit function. But in our case g is a majorization scheme for f . Thus $g(x, y) - f(x)$ has a minimum over x at y , with minimum value zero. This implies that

$$\mathcal{D}_1g(y, y) = \mathcal{D}f(y)$$

for all y , and thus, assuming f is also twice differentiable,

$$\mathcal{D}_{11}g(y, y) + \mathcal{D}_{12}g(y, y) = \mathcal{D}^2f(y),$$

Majorization also implies that $\mathcal{D}_{11}g(y, y) \succeq \mathcal{D}^2f(y)$ in the Loewner sense, i.e. $\mathcal{D}_{11}g(y, y) - \mathcal{D}^2f(y)$ is positive semi-definite. Thus $\mathcal{D}_{12}g(y, y)$ is symmetric and negative semi-definite.

Combining majorization with the derivative from the IFT shows that at a fixed point y ,

$$\mathcal{D}\omega(y) = -[\mathcal{D}_{11}g(y, y)]^{-1}\mathcal{D}_{12}g(y, y) = I - [\mathcal{D}_{11}g(y, y)]^{-1}\mathcal{D}^2f(y).$$

2.1 Example: SMACOF

The Euclidean least squares metric multidimensional scaling problem is the minimization, after a change of coordinates detailed in De Leeuw (1993), of the function

$$f(x) = 1 - \sum_{k=1}^K w_k \delta_k \sqrt{x' A_k x} + \frac{1}{2} x' x,$$

where the A_i are positive semi-definite matrices of order n that add up to the identity. The w_k and δ_k are positive **weights** and **dissimilarities**. It follows that, provided $x' A_k x > 0$ for all k , that

$$\mathcal{D}f(x) = x - B(x)x,$$

where

$$B(x) \triangleq \sum_{k=1}^K w_k \frac{\delta_k}{\sqrt{x' A_k x}} A_k$$

Also

$$\mathcal{D}^2f(x) = I - H(x),$$

where

$$H(x) \triangleq \sum_{k=1}^K w_k \frac{\delta_k}{\sqrt{x' A_k x}} \left\{ A_k - \frac{A_k x x' A_k}{x' A_k x} \right\}.$$

The SMACOF majorization algorithm (De Leeuw (1977), De Leeuw and Heiser (1977), De Leeuw and Heiser (1980), De Leeuw (1984), De Leeuw (1988), Borg and Groenen (2005), De Leeuw and Mair (2009)) uses the Cauchy-Schwartz inequality in the form

$$\sqrt{x' A_k x} \geq \frac{1}{\sqrt{y' A_k y}} x' A_k y.$$

This implies that g , defined as

$$g(x, y) = 1 - x' B(y) y + \frac{1}{2} x' x$$

majorizes f at y , which leads to the algorithm

$$x^{(k+1)} = B(x^{(k)}) x^{(k)}.$$

By direct calculation the derivatives of the algorithmic map at a fixed point y is $H(y)$. From our previous implicit function formulas we get

$$\mathcal{D}\omega(y) = I - [\mathcal{D}_{11}g(y, y)]^{-1} \mathcal{D}^2 f(y) = I - (I)^{-1} (I - H(y)) = H(y),$$

while also $\mathcal{D}_{12}g(y, y) = -H(y)$.

3 Equality Constraints

Suppose $h_j : \mathbb{R}^n \Rightarrow \mathbb{R}$ are m continuously differentiable functions, and

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid h_j(x) = 0\}.$$

We assume that for each y the minimum of g over $x \in \mathcal{X}$ is unique and attained at $\omega(y)$, and that the m vectors $\mathcal{D}h_j(\omega(y))$ are linearly independent. At a solution of the stationary equations we have

$$\begin{aligned} \mathcal{D}_1 g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}h_j(\omega(y)) &= 0, \\ h_j(\omega(y)) &= 0, \end{aligned}$$

where the λ_j are Lagrange multipliers, which are uniquely defined because of the linear independence assumption.

Define

$$A \triangleq \mathcal{D}_{11}g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}^2 h_j(\omega(y)),$$

and

$$\begin{aligned} B &\triangleq \mathcal{D}_1 h(\omega(y)), \\ E &\triangleq \mathcal{D}_{12}g(\omega(y), y), \\ F &\triangleq \mathcal{D}\omega(y), \\ G &\triangleq \mathcal{D}\lambda(y). \end{aligned}$$

Differentiating the stationary equations gives

$$\begin{bmatrix} \mathcal{D}_{11}g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}^2 h_j(\omega(y)) & \mathcal{D}_1 h(\omega(y)) \\ \mathcal{D}_1 h(\omega(y)) & 0 \end{bmatrix} \begin{bmatrix} \mathcal{D}\omega(y) \\ \mathcal{D}\lambda(y) \end{bmatrix} = \begin{bmatrix} -\mathcal{D}_{12}g(\omega(y), y) \\ 0 \end{bmatrix}$$

$$AF + B'G = -E,$$

$$BF = 0,$$

which has the solution, assuming A is invertible,

$$\begin{aligned} G &= -(BA^{-1}B')^{-1}BA^{-1}E, \\ F &= -A^{-1}(I - B'(BA^{-1}B')^{-1}BA^{-1})E. \end{aligned}$$

The convergence rate of the algorithm is the eigenvalue of F with largest modulus, evaluated at the fixed point where $\omega(y) = y$.

Also, majorization of f by g implies that $\mathcal{D}_1g(x, x) = \mathcal{D}f(x)$ for all x , and thus at a fixed point

$$-E = \mathcal{D}_{12}g(x, x) = \mathcal{D}_{11}g(x, x) - \mathcal{D}^2f(x).$$

This, in turn, implies that $-E$ is symmetric and positive semi-definite.

3.1 Example: Power Method

Consider the problem of minimizing f with $f(x) = -\frac{1}{2}x'Vx$, with V positive definite of order n , over all x satisfying $x'x = 1$ and $Ux = 0$, with U an $m \times n$ matrix of rank m .

Since $-\frac{1}{2}(x - y)'V(x - y) \leq 0$ we see that g with $g(x, y) = \frac{1}{2}y'Vy - x'Vy$ is a suitable majorization function. In this case we can derive an explicit formula for the update map, and consequently we can check if the results correspond with the ones derived from our version of the IFT for the Langrangian.

The Langrangian stationary equations for solving the majorization iterations are

$$\begin{aligned} -Vy + \lambda x + U'\mu &= 0, \\ x'x &= 1, \\ Ux &= 0. \end{aligned}$$

Thus $\mu = (UU')^{-1}UVy$ and $QVy = \lambda x$, with Q the orthogonal projector

$$Q \triangleq I - U'(UU')^{-1}U.$$

We find $\lambda = \sqrt{y'VQVy}$, and the update map is

$$\omega(y) = \frac{1}{\lambda}QVy.$$

Thus the majorization algorithm is the power method for the matrix QV , which converges to its largest eigenvalue and corresponding right eigenvector.

Taking derivatives of the update map

$$\mathcal{D}\omega(y) = \frac{1}{\lambda} \left\{ QV - \frac{1}{\lambda^2} QVyy'VQV \right\},$$

At a stationary point $QVy = \lambda y$, and thus

$$\mathcal{D}\omega(y) = \frac{1}{\lambda} \{ QV - yy'V \}$$

Note that if y is a right eigenvector of QV then Vy is a left eigenvector. The convergence rate of the majorization algorithm is the ratio of the second largest to the largest eigenvalue of QV .

In addition we find $\mathcal{D}\mu(y) = (UU')^{-1}UV$ and $\mathcal{D}\lambda(y) = \frac{1}{\lambda}VQVy$, which is $\mathcal{D}\lambda(y) = Vy$ at a stationary point.

Now let's see if we get the same solution by solving the system

$$\begin{aligned} AF + B'G &= -E, \\ BF &= 0. \end{aligned}$$

For our example

$$\begin{aligned} A &= \lambda I, \\ B &= \begin{bmatrix} x' \\ U \end{bmatrix}, \\ E &= -V, \end{aligned}$$

We must have

$$\lambda F + \begin{bmatrix} x & U' \end{bmatrix} G = V,$$

and

$$\begin{bmatrix} x' \\ U \end{bmatrix} F = 0.$$

For the derivatives of the Lagrange multipliers we find

$$G = \begin{bmatrix} x'V \\ (UU')^{-1}UV \end{bmatrix},$$

which corresponds with our results using the explicit formulas. Also

$$\lambda F = V - \begin{bmatrix} x & U' \end{bmatrix} G = V - xx'V - U(UU')^{-1}UV = QV - xx'V,$$

which is indeed what we found using the explicit expression for ω .

3.2 Example: Least Squares with Quadratic Constraint

Consider the problem of minimizing

$$f(x) = \frac{1}{2}(x - u)'V(x - u)$$

over all x satisfying $\frac{1}{2}x'Wx = 1$. Here W is assumed to be positive definite. This problem, or some variation of it, has been studied by many authors, starting with Forsythe and Golub (1965), Spjøtvoll (1972), Gander (1981).

We use majorization to construct an iterative algorithm. Suppose κ is such that $V \preceq \kappa W$. Then

$$f(x) \leq f(y) + (x - y)'V(y - z) + \frac{1}{2}\kappa(x - y)'W(x - y).$$

We can minimize the majorization function over $\frac{1}{2}x'Wx = 1$ by minimizing the linear function $x'V(y - z) - \kappa x'Wy$ over $\frac{1}{2}x'Wx = 1$. The minimum is attained for

$$x \propto y - \frac{1}{\kappa}W^{-1}V(y - z),$$

and we find the update by normalizing. The derivative of the update function, computed directly, is

$$\mathcal{D}\omega(y) = \left[I - \frac{yy'W}{y'Wy} \right] \left[I - \frac{1}{\kappa}W^{-1}V \right]$$

The Lagrangian equations for the majorization problem are

$$\begin{aligned} V(y - z) - \kappa Wy + \lambda(y)W\omega(y) &= 0, \\ \frac{1}{2}\omega(y)'W\omega(y) &= 1 \end{aligned}$$

Thus

$$\begin{aligned} \lambda'(y)W\omega(y) + \lambda(y)W\mathcal{D}\omega(y) &= 0, \\ \omega(y)'W\mathcal{D}\omega(y) &= 0. \end{aligned}$$

$$\mathcal{D}\omega(y) = -\frac{\lambda'(y)}{\lambda(y)}\omega(y)$$

Thus $\lambda'(y) = \omega(y)'W\omega(y)$ and $\lambda(y) = \mathcal{D}\omega(y)'W\mathcal{D}\omega(y)$.

The IFT with

$$\begin{aligned} A &= \lambda W \\ B &= Wx, \\ E &= V - \kappa W \end{aligned}$$

$$F = -(\lambda W)^{-1}(I - x'W(Wx(\lambda W)^{-1}xW)^{-1}Wx(\lambda W)^{-1})(V - \kappa W) = -\lambda^{-2}W^{-1}\left(I - \frac{\lambda}{x'W^{-1}x}xx'W\right)^{-1}x'W^{-1}$$

4 Parametric Equality Constraints

A more general problem is to compute, and study,

$$\omega(y) \triangleq \mathbf{argmin} \{g(x, y) \mid H(x, y) = 0\},$$

$$\mathcal{X}(y) = \{x \in \mathbb{R}^n \mid \bigvee_{j=1}^m h_j(x, y) = 0\}.$$

where the constraints also depend on the parameter y . The relevance for majorization algorithms may not be immediately obvious. When treating inequality constraints in a subsequent section, however, the formulas in this section will be useful.

The Lagrangian equations are

$$\begin{aligned} \mathcal{D}_1 g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}_1 h_j(\omega(y), y) &= 0, \\ H(\omega(y), y) &= 0. \end{aligned}$$

Differentiating again gives

$$\begin{aligned} \mathcal{D}_{11} g(\omega(y), y) \mathcal{D}x(y) + \mathcal{D}_{12} g(\omega(y), y) + \\ \sum_{j=1}^m \lambda_j(y) \{ \mathcal{D}_{11} h_j(\omega(y), y) \mathcal{D}\omega(y) + \mathcal{D}_{12} h_j(\omega(y), y) \} + \\ \sum_{j=1}^m \mathcal{D}\lambda_j(y) \mathcal{D}_1 h_j(\omega(y), y) &= 0, \end{aligned}$$

and

$$\mathcal{D}_1 h_j(\omega(y), y) \mathcal{D}\omega(y) + \mathcal{D}_2 h_j(\omega(y), y) = 0.$$

Define

$$A \triangleq \mathcal{D}_{11} g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}_{11} h_j(\omega(y), y),$$

and

$$\begin{aligned} B &\triangleq \mathcal{D}_1 h(\omega(y), y), \\ C &\triangleq \mathcal{D}_2 h(\omega(y), y), \\ D &\triangleq \sum_{j=1}^m \lambda_j(y) \mathcal{D}_{12} h_j(\omega(y), y), \\ E &\triangleq \mathcal{D}_{12} g(\omega(y), y), \\ F &\triangleq \mathcal{D}\omega(y), \\ G &\triangleq \mathcal{D}\lambda(y). \end{aligned}$$

We must solve

$$\begin{aligned} AF + B'G &= -(E + D), \\ BF &= -C. \end{aligned}$$

We have $F = -A^{-1}(B'G + E + D)$ and $BA^{-1}(B'G + E + D) = C$. Thus

$$G = (BA^{-1}B')^{-1}\{C - BA^{-1}(E + D)\},$$

and

$$F = -A^{-1}B'(BA^{-1}B')^{-1}C - \{A^{-1} - A^{-1}B'(BA^{-1}B')^{-1}BA^{-1}\}(E + D).$$

We recover the results for the basic case from the previous section by setting $C = 0$ and $D = 0$.

5 Inequality Constraints, Part I

If the minimization problem solved in the majorization step involves inequality constraints

$$\omega(y) \triangleq \mathbf{argmin} \{g(x, y) \mid H(x) = 0 \wedge K(x) \geq 0\},$$

$$\mathcal{X}(y) = \{x \in \mathbb{R}^n \mid \bigvee_{j=1}^{\ell} h_j(x, y) = 0, \bigvee_{j=\ell+1}^m h_j(x, y) \geq 0\}.$$

then we cannot expect differentiability any more.

Of course an inequality constraint like $h_j(x) \geq 0$ can be rewritten as an equality constraint $h_j(x) - z_j^2 = 0$, where the z_j are additional variables.

Just consider the very simple example

$$\theta(y) = \min \{(x - y)^2 \mid x \geq 0\} = \begin{cases} 0 & \text{if } y \geq 0, \\ y^2 & \text{otherwise,} \end{cases}$$

and

$$\omega(y) = \mathbf{argmin} \{(x - y)^2 \mid x \geq 0\} = \begin{cases} y & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The value function θ is differentiable at zero, but not twice differentiable. The solution function ω is continuous, but not differentiable at zero. At zero it does, however, have a left derivative equal to zero and a right derivative equal to one.

Minimizing $g(x, y) = (x^2 - y^2)^2$ over $x \geq 0$ gives θ identically equal to zero, and $\omega(y) = |y|$, continuous but not differentiable at zero. For $g(x, y) = (x - y^2)^2$ and $x \geq 0$ we again find θ identically equal to zero but now $\omega(y) = y^2$, which is differentiable. These small examples show that in general the solution function will not be differentiable, but may be directionally differentiable.

Convert

$$k_j(x) - z_j^2 = 0$$

If our original problem is to minimize f over all x satisfying $h_j(x) = 0$ and $p_\ell(x) \leq 0$, then it may make sense to majorize both f and the k_ℓ .

If $g : \mathbb{R} \otimes \mathbb{R} \Rightarrow \mathbb{R}$ majorizes f at y and $q_\ell : \mathbb{R} \otimes \mathbb{R} \Rightarrow \mathbb{R}$ majorizes k_ℓ at y , then the majorization algorithm is

$$x^{(k+1)} = \mathbf{argmin} \{g(x, x^{(k)}) \mid h_j(x) = 0 \wedge q_\ell(x, y^{(k)}) \leq 0\}$$

In fact we will look at the more general problem

$$\omega(y) = \mathbf{argmin}_x \{g(x, y) \mid h_j(x, y) = 0 \wedge p_\ell(x, y) \leq 0\}$$

From the Karush-Kuhn-Tucker necessary conditions for this minimization problem

$$\mathcal{D}_1 g(\omega(y), y) + \sum_{j=1}^m \lambda_j(y) \mathcal{D}_1 h_j(\omega(y), y) + \sum_{\ell=1}^r \gamma_\ell(y) \mathcal{D}_1 p_\ell(\omega(y), y) = 0$$

$$\begin{aligned} h_j(\omega(y), y) &= 0, \\ p_\ell(\omega(y), y) &\leq 0, \\ \gamma_\ell(y) &\geq 0, \\ \gamma_\ell(y) p_\ell(\omega(y), y) &= 0. \end{aligned}$$

6 Inequality Constraints, Part II

If the conditions of the previous section are not satisfied we cannot guarantee differentiability, and we need to take a step back. First, we need to use the version of the linear convergence theorem for fixed point iterations from Sheng and Xu (2001). Then we need to step down a level from continuous differentiability to directional differentiability and the Clark Jacobian, following Dempe and Pallaschke (1997).

References

- Bank, B., J. Guddat, D. Klatte, B. Kummer, and K. Tammer. 1982. *Non-Linear Parametric Optimization*. Akademie-Verlag.
- Bonnans, J.F., and A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer.
- Borg, I., and P.J.F. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications*. Second Edition. Springer.

- De Leeuw, J. 1977. “Applications of Convex Analysis to Multidimensional Scaling.” In *Recent Developments in Statistics*, edited by J.R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem, 133–45. Amsterdam, The Netherlands: North Holland Publishing Company. http://www.stat.ucla.edu/~deleeuw/janspubs/1977/chapters/deleeuw_C_77.pdf.
- . 1984. “Differentiability of Kruskal’s Stress at a Local Minimum.” *Psychometrika* 49: 111–13. http://www.stat.ucla.edu/~deleeuw/janspubs/1984/articles/deleeuw_A_84f.pdf.
- . 1988. “Convergence of the Majorization Method for Multidimensional Scaling.” *Journal of Classification* 5: 163–80. http://www.stat.ucla.edu/~deleeuw/janspubs/1988/articles/deleeuw_A_88b.pdf.
- . 1993. “Fitting Distances by Least Squares.” Preprint Series 130. Los Angeles, CA: UCLA Department of Statistics. http://www.stat.ucla.edu/~deleeuw/janspubs/1993/reports/deleeuw_R_93c.pdf.
- . 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. http://www.stat.ucla.edu/~deleeuw/janspubs/1994/chapters/deleeuw_C_94c.pdf.
- . 2016. *Block Relaxation Methods in Statistics*. Bookdown. <https://bookdown.org/jandeleeuw6/bras/>.
- De Leeuw, J., and W.J. Heiser. 1977. “Convergence of Correction Matrix Algorithms for Multidimensional Scaling.” In *Geometric Representations of Relational Data*, edited by J.C. Lingoes, 735–53. Ann Arbor, Michigan: Mathesis Press. http://www.stat.ucla.edu/~deleeuw/janspubs/1977/chapters/deleeuw_heiser_C_77.pdf.
- . 1980. “Multidimensional Scaling with Restrictions on the Configuration.” In *Multivariate Analysis, Volume V*, edited by P.R. Krishnaiah, 501–22. Amsterdam, The Netherlands: North Holland Publishing Company. http://www.stat.ucla.edu/~deleeuw/janspubs/1980/chapters/deleeuw_heiser_C_80.pdf.
- De Leeuw, J., and P. Mair. 2009. “Multidimensional Scaling Using Majorization: SMACOF in R.” *Journal of Statistical Software* 31 (3): 1–30. http://www.stat.ucla.edu/~deleeuw/janspubs/2009/articles/deleeuw_mair_A_09c.pdf.
- Dempe, S., and D. Pallaschke. 1997. “Quasidifferentiability of Optimal Solutions in Parametric Nonlinear Optimization.” *Optimization* 40 (1): 1–24.
- Dontchev, A.L., and R.T. Rockafellar. 2014. *Implicit Functions and Solution Mappings*. Second Edition. Springer Series in Operations Research and Financial Engineering. Springer.
- Fiacco, A.V. 1983. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press.
- Forsythe, G.E., and G.H. Golub. 1965. “On the Stationary Values of a Second Degree Polynomial on the Unit Sphere.” *Journal of the Society for Industrial and Applied Mathematics* 13: 1050–68.

- Gander, W. 1981. “Least Squares with a Quadratic Constraint.” *Numerische Mathematik* 36: 291–307.
- Giorgi, G., and C. Zuccotti. 1918. “A Tutorial on Sensitivity and Stability in Nonlinear Programming and Variational Inequalities under Differentiability Assumptions.” DEM Working Paper Series 159 (06-18). Department of Economics; Management, Università di Pavia. <http://economieaweb.unipv.it/wp-content/uploads/2018/01/DEMWP0159.pdf>.
- Heiser, W.J. 1995. “Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis.” In *Recent Advances in Descriptive Multivariate Analysis*, edited by W.J. Krzanowski, 157–89. Oxford: Clarendon Press.
- Lange, K. 2016. *MM Optimization Algorithms*. SIAM.
- Levitin, E.S. 1994. *Perturbation Theory in Mathematical Programming and Its Applications*. Wiley.
- Luderer, B., L. Minchenko, and T. Satsura. 2002. *Multivalued Analysis and Nonlinear Programming Problems with Perturbations*. Vol. 66. Nonconvex Optimization and Its Applications. Kluwer Academic Publishers.
- Sheng, S., and H. Xu. 2001. “Picard Iteration for Nonsmooth Equations.” *Journal of Computational Mathematics* 19 (6): 583–90.
- Spjøtvoll, E. 1972. “A Note on a Theorem of Forsythe and Golub.” *SIAM Journal on Applied Mathematics* 23: 307–11.
- Still, G.J. 2018. “Lectures on Parametric Optimization: An Introduction.” http://www.optimization-online.org/DB_FILE/2018/04/6587.pdf.