ANALISIS 9 (1 & 2), 63-76 (2002)

SCALING SYMMETRIC RANK ONE UPDATE FOR UNCONSTRAINED OPTIMIZATION

MALIK ABU HASSAN
MANSOR MONSI
LEONG WAH JUNE
Department of Mathematics
Universiti Putra Malaysia

ABSTRACT

A basic disadvantage to the symmetric rank one (SR1) update is that the SR1 update may not preserve positive definiteness when starting with a positive definite approximation. A simple remedy to this problem is to restart the update with the initial approximation mostly the identity matrix whenever this difficulty arises. However, numerical experience shows that restart with the identity matrix is not a good choice. Instead of using the identity matrix we used a positive multiple of the identity matrix. The used positive scaling factor is the optimal solution of the measure defined by the problem – maximize the determinant subject to a bound of 1 on the largest eigenvalue. This measure is motivated by considering the volume of the symmetric difference of the two ellipsoids, which arise from the current and updated quadratic models in quasi-Newton methods. A replacement in the form of positive multiple of identity matrix is provided for the SR1 when it is not positive definite. Our experiments indicate that with such simple scale, the effectiveness of the SR1 method is increased dramatically.

Keywords. *Symmetric rank one, Volume of ellipsoid, Unconstrained optimization.*

ABSTRAK

Kelemahan asas bagi pengemaskini pangkat satu yang simetri (SR1) ialah pengemaskini SR1 tidak dapat mengekal ketentu positif bila bermula dengan suatu hampiran tentu positif. Penanganan yang mudah terhadap masalah tersebut ialah memula-semula pengemaskini dengan hampiran awal yang biasanya merupakan matriks identiti apabila masalah tersebut timbul. Walau bagaimanapun, pengalaman berangka menunjukkan bahawa bermula-semula dengan matriks identiti bukanlah suatu pilihan yang baik. Gandaan positif

matriks identiti digunakan untuk menggantikan matriks identiti. Faktor penskalaan yang digunakan itu merupakan penyelesaian optimum bagi ukuran yang ditakrifkan oleh masalah-memaksimumkan penentu tertakluk kepada pembatasan 1 atas nilaieigen terbesar. Ukuran tersebut dimotivasikan dengan mempertimbangkan isipadu beza simetri bagi dua ellipsoid yang timbul dari model kuadratik kemaskini dan semasa dalam kaedah kuasi-Newton. Jadi, suatu penggantian dalam bentuk gandaan positif matriks identiti telah dibekalkan untuk SR1 apabila ia tidak tentu positif. Ujikaji kami menunjukkan bahawa dengan skala mudah seperti itu, kecekapan kaedah SR1 meningkat dengan berkesan sekali.

Katakunci. *Pangkat satu yang simetri, isipadu ellipsoid, pengoptimuman tak berkekangan.*

INTRODUCTION

This paper is concerned with quasi-Newton methods for finding a local minimum of the unconstrained optimization problem,

$$\min f(x) \qquad (1.1)$$

$$x \in \Re^n$$

It will be assumed that $f(x)$ is at least twice continuously differentiable. Algorithms for solving (1.1) are iterative and the basic framework of an iteration of a secant method is:
Given current iteration x_c $f(x_c)$ $\nabla f(x_c)$ or finite difference approximation, and or finite difference approximation, and $B_c \in \Re^{n\times n}$ symmetric (secant approximation to $\nabla^2 f(x_c)$ select new iterate x_+ by a line search method. Update B_c to B_+ such that B_+ is symmetric and satisfies the secant equation $B_+ s_c = y_c$, where $s_c = x_+ - x_c$ and $y_c = \nabla f(x_+) - \nabla f(x_c)$.

In this paper, we consider the SR1 update for the Hessian approximation,

$$B_+ = B_c + \frac{(y_c - B_c s_c)(y_c - B_c s_c)^{\mathrm{T}}}{s_c^{\mathrm{T}}(y_c - B_c s_c)} \qquad (1.2)$$

and throughout if $H = B^{-1}$, the inverse update respected to SR1 is given by

$$H_+ = H_c + \frac{(s_c - H_c y_c)(s_c - H_c y_c)^T}{y_c^T (s_c - H_c y_c)} \tag{1.3}$$

For the background on these updates and others see Fletcher (1980), Gill et al. (1981), and Dennis and Schnabel (1983).

The SR1 formula makes a symmetric rank one change to the previous Hessian approximation B_c.Compared with other secant updates, the SR1 update is simpler and may require less computation per iteration when unfactored forms of updates are used. (Factored updates are those in which a decomposition of B_c is updated at each iteration.) The SR1 update has a major drawback in that it does not guarantee positive definiteness. However, it has some very strong convergence properties. Under certain regularity conditions, the updates converge globally to the true Hessian. Successful numerical tests— in a trust region framework to avoid the possible loss of positive definiteness — has resulted in a renewed interest in the SR1 update, see e.g., Khalfan (1989). Another method of avoiding the loss of positive definiteness of the SR1 is to size the current update, see IP and Todd (1988). The resulting updates are called the optimal conditioned sized SR1 updates.

The primary motivation for this paper is to find another method of avoiding the loss of positive definiteness of the SR1. In the next section, we present a simple restart procedure for the SR1 method using standard line search to avoid the loss of positive definiteness of the SR1.

RESTART PROCEDURE FOR SR1 UPDATE

In this section, we present an algorithm using SR1 update and a restart procedure for unconstrained minimization.

Algorithm 2.1. Quasi-Newton SR1 method with Restart (NSSR1)
Step 0. Given an initial point x_0, an initial positive matrix $H_0 = I$, set $k = 0$.
Step 1. If the convergence criterion

$$\|\nabla f(x_k)\| \le \varepsilon \times \max(1, \|x_k\|)$$

is achieved, the stop.

Step 2. Compute a quasi-Newton direction

$$p_k = -H_k \nabla f(x_k) \tag{2.2}$$

where H_K is given by (1.3).

Step 3. If $p_k^{\mathrm{T}} \nabla f(x_k) > 0$, ($H_k$ is not positive definite) set $H_k = I$ and subsequently

$p_k = -\nabla f(x_k)$. Else retain (2.2).

Step 4. Using a backtracking line search, find an acceptable steplength λ_k such that the Wolfe's (1969) condition

$$f(x_k + \lambda_k p_k) \leq f(x_k) + \alpha \lambda_k \nabla f(x_k)^{\mathrm{T}} p_k \tag{2.3}$$

and

$$\nabla f(x_k + \lambda_k p_k)^{\mathrm{T}} p_k \geq \alpha' \nabla f(x_k)^{\mathrm{T}} p_k \tag{2.4}$$

is satisfied. $\lambda_k = 1$ is always tried first $\alpha = 10^{-4}$ and $\alpha' = 0.9$).

Step 5. Set $x_{k+1} = x_k + \lambda_k p_k$.

Setp 6. Compute the next inverse Hessian approximation H_{k+1}.

Step 7. Set $k = k + 1$, and go to Step 1.

TOMS 500 unconstrained optimization code (Shanno and Phua, 1980) is modified to fit with Algorithm 2.1 (NSSR1) where BFGS update in the code is replaced by SR1 update and a restart procedure is provided. The stopping tolerance, ε used was 10^{-5}. The restart procedure provides a replacement for the non-positive definite H_k with the identity matrix. The problems tested are: Penalty function I, II, Rosenbrook function, Powell function, Wood function, Beale function (Moré et al., 1981) and Trigonometric function (Conn et al., 1988). The results are given in Table 1.

In the table, n denotes the number of variables, n_i the number of iterations, n_f the number of function evaluations and $n_{restart}$ the number of restart. For all methods the number of gradient evaluations equals the number of function evaluations. The algorithm also stopped when the number of function evaluation exceeds 999 and the symbol "*EX*" is used.

Table 1
Results for NSSR1 Algorithm

	n_1	n_f	$n_{restart}$
Penalty I			
$n = 4$	85	169	15
$n = 20$	55	99	9
$n = 100$	79	159	11
$n = 400$	68	60	8
Penalty II			
$n = 4$	21	43	1
$n = 20$	EX	EX	EX
$n = 100$	EX	EX	EX
$n = 400$	EX	EX	EX
Trigonometric			
$n = 4$	16	25	1
$n = 20$	57	101	21
$n = 100$	60	85	28
$n = 400$	68	93	33
Rosenbrook			
$n = 4$	60	101	13
$n = 20$	EX	EX	EX
$n = 100$	EX	EX	EX
$n = 400$	EX	EX	EX
Powell			
$n = 4$	34	49	1
$n = 20$	EX	EX	EX
$n = 100$	EX	EX	EX
$n = 400$	EX	EX	EX
Wood			
$n = 4$	110	234	22
$n = 20$	EX	EX	EX
$n = 100$	EX	EX	EX
$n = 400$	EX	EX	EX
Beale			
$n = 4$	15	25	2
$n = 20$	13	19	1
$n = 100$	12	22	0
$n = 400$	13	23	1

The numerical results in Table 1 show that restarting with the identity may be very unsuitable. The SR1 update may not preserve positive definiteness at the next iteration even if the current does, i.e., when H_c is positive definite and $s_c^T y_c > 0$. The algorithm will keep on restarting with little or no progress until the maximum number of function evaluation allowed is exceeded. Therefore, restart with the identity matrix is clearly not a good choice. Instead, we consider the cheap choice of replacing the identity as a positive multiple of the identity matrix. In the following section, a positive scale is derived.

SCALING THE IDENTITY

Preliminaries

In the following, we will use the notation that

$$a = y^T H_c y\,,\ b = y^T s\,,\ c = s^T B_c s\,. \tag{3.1}$$

Throughout the section we deal with the space of real symmetric $n \times n$ matrices, n ≥ 2, equipped with the trace inner product, $\langle A, B\rangle$, = trace (AB) and the induced Frobenius norm, $||A||_F = \langle A,A\rangle^{1/2}$. We assume the curvature condition $b = y^T s > 0$ and that the current Hessian approximation B_C is positive definite.

The primary motivation for this paper is to find the 'best' new update B_+, i.e. this update should satisfy the secant equation while preserving the most information from the current update B_c. With this aim in mind, we first show that minimizing the volume of the symmetric difference between the two ellipsoids corresponding to B_+, and B_c is a valid measure for preserving the most information. This hard problem is not solved but rather relaxed in several ways. This leads to the measures that yield SR1 updates. Adding the restriction that the ellipsoid for B_+ contains (or is contained in) the normalized ellipsoid for B_c yields the measure

$$\sigma(A) = \frac{\lambda_1(A)}{\det(A)^{1/n}} \tag{3.2}$$

where det denotes determinant and A is chosen to be scaled updates $H_c^{1/2} B^{+} H_c^{1/2}$ and $B_c^{1/2} H_{+} B_c^{1/2}$, respectively. λ_1 is the largest eigenvalue of A

The optimal updates for this measure are the optimal conditioned, sized, SR1 updates. The sizing factor for the ellipsoids corresponding to B_{+} is useful for us to derive our scaling factor. In the next sub-section, we present several results due to Wolkowicz (1993) on volumes of ellipsoid, which lead to the measure σ.

Volume as a Measure for Least Change

In this sub-section we derive a measure of least change. This measure arises from relaxation of the problem: approximate a given ellipsoid by another ellipsoid, from within a given set, by minimizing the volume of their symmetric difference. This measure involves the singular values of the product of two semi positive definite (s.p.d) matrices. Further relaxation results in more tractable measures involve eigenvalues.

Least change secant methods attempt to find an update B_{+} that satisfies the secant equation while simultaneously preserving as much information as possible from the current Hessian approximation B_c. If we assume that the gradient vector $\nabla f(x_{+})$ can be a random direction (of norm 1 say), then we can consider that B_{+} is preserving the information from B_c when the search directions $H_{+} \nabla f(x_{+})$ and $H_c \nabla f(x_{+})$ are close. Thus B_{+} is a least change update of B_c if the ellipsoids formed from the images of the unit ball under H_{+} and H_c are closed. Let us now use the volume as a measure of closeness for ellipsoids. It would be best if we could find the update H_{+} so that the volume of the symmetric difference (set union minus intersection) of the updated and current ellipsoids is minimized. With this aim in mind, we first consider two 'optimal' updated ellipsoids. The first ellipsoid minimizes the volume over all ellipsoid, while the second one maximize the volume over all ellipsoids contained within the current ellipsoid.

Support that B is s.p.d Denote the ellipsoid for B of radius α by

$$E_\alpha(B) = \{x \in \Re^n : \|Bx\| \le \alpha\} = \{x \in \Re^n : x^{\mathrm{T}} B^2 x \le \alpha^2\} \; ; (3.3)$$

Denote the ellipsoid corresponding to the square root of B by

$$E_\alpha(B^{1/2}) = \{x \in \Re^n : x^{\mathrm{T}} B x \le \alpha^2\} . \qquad (3.4)$$

Note that the image of the ball under H is $H(E_\alpha(I))=E_\alpha(B)$. The volume of this ellipsoid is the determinant of H times the volume of $E_\alpha(I)$, i.e.

$$vol(E_\alpha(B)) = \frac{\alpha^n}{\det(B)} vol(E_\alpha(I)) \cdot$$

Given B fixed and, since

$$\lambda_1(B) = \max_{x \in E_\alpha(I)} \|Bx\|,$$

the ellipsoid of minimal volume containing $E_\alpha(I)$ is $E_\alpha(B)$ with $\alpha = \lambda_1(B)$. The n-th root of the volume of this ellipsoid leads to our measure

$$\sigma(B) = \frac{\lambda_1(B)}{\det(B)^{1/n}}$$

The measure σ has several interesting properties.

Proposition 3.1. If $\kappa(B) = \dfrac{\lambda_1(B)}{\lambda_n(B)}$ (the ℓ_2 condition number)

where λ_1 and λ_n are the largest and smallest eigenvalues respectively, the σ (B) satisfies

1. $1 \le \sigma(B) \le n\kappa(B) \le 4n\sigma^n(B)$

2. $\sigma(\alpha B) = \sigma(B)$ for all $\alpha > 0$;

3. σ is a pseudoconvex function on the set of s.p.d. matrices and thus any stationary point is a global minimizer.

Proof. See Wolkowicz (1993).

The inequalities in 1. of the Proposition show that σ acts as condition number in the sense that it provides bounds on the amplification factors for relative error. Moreover, since the σ bounds κ from below and above, minimizing one would be a compromise for minimizing the other.

The σ- Optimal Update

We now show that the best s.p.d. updates for the measure σ are sized, optimally conditioned, SR1 updates in IP and Todd (1988). Thus these updates provide ellipsoids of minimum (maximum) volume containing (contained in) the current normalized ellipsoid. We again assume that $b > 0$ and B_c is s.p.d.

Theorem 3.1. Let

$$\delta_{\pm} = \frac{c}{b} \pm \{\frac{c^2}{b^2} - \frac{c}{a}\}^{1/2} \tag{3.5}$$

Then the SR1 update of $\frac{1}{\delta} B_c$,

$$H_+ = \delta H_c + vv^{\mathrm{T}} / (v^{\mathrm{T}} y), \text{where } v = s - \delta H_c y,\ \delta = \delta_-, \tag{3.6}$$

is the unique solution of

$$\min \sigma (H_c B_+)$$

$$\text{s.t.} \qquad B_+ s = y,\ B_+ \text{ is s.p.d..}$$

Moreover, $1/\delta = \lambda_1 (H_c B_+)$ is multiplicity n - 1 and the other eigenvalue of $H_c B_+$ is $\lambda n (H_c B_+) = 1/\delta_+$.

Proof. The Theorem and its proof can be found in Wolkowicz (1993).

Corollary 3.1. Let

$$\hat{\delta}_{\pm} = \frac{a}{b} \pm \{\frac{a^2}{b^2} - \frac{a}{c}\}^{1/2} \tag{3.7}$$

Then the SR1 update of $\frac{1}{\hat{\delta}} H_c$,

$$B_+ = \hat{\delta} B_c + \hat{v}\hat{v}^{\mathrm{T}} / (\hat{v}^{\mathrm{T}} s), \text{where } \hat{v} = y - \hat{\delta} B_c s,\ \hat{\delta} = \hat{\delta}_-, \tag{3.8}$$

is the unique solution of

$$\min \sigma(H_c B_+)$$
$$\text{s.t.} \quad B_+ s = y, \; B_+ \text{ is s.p.d..}$$

Moreover, $1/\hat{\delta} = \lambda_1(B_c H_+)$ is of multiplicity n - 1 and equals δ_+ in (3.5); the other eigenvalue of $B_c H_+$ is $1/\hat{\delta}_+$ and equals δ_- the reciprocal from (3.5); the largest and smallest eigenvalues of $H_c B_+$ and the optimal value of the measure, from Theorem 3.1 and the Corollary, all have the same respective values.

Proof. The proof follows by interchanging the roles of H and B. It can be seen that the optimal values are the same for both problems by using the fact that largest n-1 eigenvalues are equal at the optimum in Theorem 3.1 while the smallest n - 1 eigenvalues are equal in the Corollary and $\kappa(B) = \kappa(B^{-1})$.

We can now give our optimal scaling factor under the measure σ for SR1 update:

Theorem 3.2. Let $\hat{a} = y^T y$, $\hat{c} = s^T s$ and

$$\tilde{\delta} = \frac{\hat{c}}{b} - \{\frac{\hat{c}^2}{b^2} - \frac{\hat{c}}{\hat{a}}\}^{1/2} \tag{3.9}$$

Then the SR1 update of,

$$H_+ = \tilde{\delta} I + vv^T/(v^T y), \; v = s - \tilde{\delta} y, \tag{3.10}$$

is the unique solution of

$$\min \sigma(B_+)$$
$$\text{s.t.} \quad B_+ s = y, \; B_+ \text{ is s.p.d..}$$

Proof. The Theorem is equivalent to Theorem 3.1 with $H_c = I$, which is the case after restarting the SR1 update.

In fact, Dennis and Wolkowicz (1990) and Wolkowicz (1993) had shown that the σ- optimal updates in Sub-Section 3.3 is actually κ- optimal as well and have a common spectral property. The κ- measure is used by Shanno and Phua (1978) to derive the optimal scaling factor for the BFGS update.

SCALED SR1 UPDATE

We now present a description of the scaled SR1 (SSR1) algorithm that ensures the positive definiteness of SR1 update.

Algorithm 4.1. Algorithm SSR1

Step 0. Given an initial point x_0, an initial positive matrix $H_0 = I$, set $k = 0$.

Step 1. If the convergence criterion

$$\|\nabla f(x_k)\| \leq \varepsilon \times \max(1, \|x_k\|) \tag{4.1}$$

is achieved, then stop.

Step 2. Compute a quasi-Newton direction

$$p_k = -H_k \nabla f(x_k), \text{ where } H_k \text{ is given by (1.3).} \tag{4.2}$$

Step 3. If $p_k^{\mathrm{T}} \nabla f(x_k) > 0$ H_k is not positive definite) or $k = 1$ set

$$H_k = \tilde{\delta}_{k-1} I$$

$$\tilde{\delta}_{k-1} = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{y_{k-1}^{\mathrm{T}} s_{k-1}} - \left\{ \frac{(s_{k-1}^{\mathrm{T}} s_{k-1})^2}{(y_{k-1}^{\mathrm{T}} s_{k-1})^2} - \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{y_{k-1}^{\mathrm{T}} y_{k-1}} \right\}^{1/2} \tag{4.3}$$

and subsequently

$$p_k = -\tilde{\delta}_{k-1} \nabla f(x_k).$$ Else retain (4.2)

Step 4. Using a backtracking line search, find an acceptable steplength, λ_k such that the Wolfe's condition (2.3)-(2.4) is satisfied. ($\lambda_k = 1$ is always tried first, $\alpha = 10^{-4}$ and $\alpha' = 0.9$).

Step 5. Set $x_{k+1} = x_k + \lambda_k p_k$.

Step 6. Compute the next inverse Hessian approximation H_{k+1}.

Step 7. Set $k = k+1$, and go to Step 1.

Algorithm SSR1 differs from NSSR1 in that the algorithm always start/restart with $\widetilde{\delta}_k I$ instead of I.

Our tests were made in double-precision arithmetic, for which the unit roundoff is approximately 10^{-16}. We compared the SSR1 algorithm with the NSSR1 algorithm described as Algorithm 2.1. The initial Hessian approximation was always the identity matrix, and after one iteration was completed, the method updated $\widetilde{\delta}_0 I$ instead of I, where $\widetilde{\delta}_0$ defined by (4.3) with k =1 for SSR1 algorithm. The results are reported in Table 2.

Table 2

Comparison of SSR1 with NSSR1

	SSR1			NSSR1		
	n_I	n_f	$n_{restart}$	n_I	n_f	$n_{restart}$
Penalty I						
$n = 4$	39	57	4	85	169	15
$n = 20$	47	80	4	55	99	9
$n = 100$	53	78	7	79	159	11
$n = 400$	60	82	3	68	60	8
Penalty II						
$n = 4$	27	30	2	21	43	1
$n = 20$	212	325	28	*EX*	*EX*	*EX*
$n = 100$	450	553	17	*EX*	*EX*	*EX*
$n = 400$	*EX*	*EX*	*EX*	*EX*	*EX*	*EX*
Trigonometric						
$n = 4$	14	21	1	16	25	1
$n = 20$	61	88	9	57	101	21
$n = 100$	56	84	15	60	85	28
$n = 400$	75	117	21	68	93	33
Rosenbrook						
$n = 4$	39	84	8	60	101	13
$n = 20$	82	132	18	*EX*	*EX*	*EX*
$n = 100$	43	63	2	*EX*	*EX*	*EX*
$n = 400$	62	89	11	*EX*	*EX*	*EX*
Powell						
$n = 4$	27	30	0	34	49	1
$n = 20$	27	30	0	*EX*	*EX*	*EX*
$n = 100$	31	35	2	*EX*	*EX*	*EX*
$n = 400$	33	40	3	*EX*	*EX*	*EX*

(*continued*)

Wood						
$n = 4$	26	35	0	110	234	22
$n = 20$	35	52	3	*EX*	*EX*	*EX*
$n = 100$	30	48	2	*EX*	*EX*	*EX*
$n = 400$	61	84	13	*EX*	*EX*	*EX*
Beale						
$n = 4$	16	21	1	15	25	2
$n = 20$	18	27	2	13	19	1
$n = 100$	19	22	1	12	22	0
$n = 400$	14	18	0	13	23	1

The results indicate that the SSR1 method is more effective than the NSSR1 in solving the given problems. We see that the SSR1 method generally requires fewer iteration and function calls than the method of NSSR1. The lack of convergence of the NSSR1 to the correct minimizer in many of these problems within a given limit of function calls can be observed in Table 1. This problem does not exist for the SSR1 method.

CONCLUSIONS

We have derived the SSR1 algorithm by choosing $\tilde{\delta}_k I$ instead I when started/restarted. Compared with the NSSR1, SR1 update requires less iterations and function calls than NSSR1 update. Moreover, we see that most of the problems can be solved by SSR1 under a certain number of function calls but not for NSSR1. Therefore, we conclude that by a simple scale on SR1 method, it can improve the SR1 method dramatically.

REFERENCES

Conn, A.R., Gould, N.I.M. & Toint, Ph. (1988). Testing a class of methods for solving minimization problems with simple bounds on the variables. *Math. Comp.* 50/2, 399-430.

Dennis, J.E. & Schnabel, R.B. (1983). *Numerical Methods for Nonlinear Equations and Unconstrained Optimization.* Englewood Cliffs, New Jersey: Prentice Hall.

Dennis, J.E. & Wolkowicz, H. (1990). *Sizing and least change secant methods* (Technical Report CORR 90-02). Waterloo, Ontario: Department of Combinatorics and Optimization, University of Waterloo.

Fletcher, R. (1980). *Practical Methods of Optimization*. New York: John Wiley & Sons.

Gill, P.E., Murray, W., & Wright, M.H. (1981). *Practical Optimization*. London: Academic Press.

IP, C.M. & Todd, M.J. (1988). Optimal conditioning and convergence in rank one quasi-Newton updates. *SIAM Journal Numerical Analysis*, 25, 206-221.

Khalfan, H.F.H. (1989). *Topics in quasi-Newton methods for unconstrained Optimization*. Unpublished doctoral dissertation, University of Colorado, Colorado.

Moré, J.J., Garbow, B.S. & Hillstrom, K.E. (1981). Testing unconstrained optimization software. *TOMS*, 7, 17-41.

Shanno, D.F. & Phua, K.H (1980). Remark on algorithm 500: minimization of unconstrained multivariate functions. *ACM Transactions on Mathematical Software*, 6, 618-622.

Shanno, D.F. & Phua, K.H. (1978). Matrix conditioning and nonlinear optimization. *Mathematical Programming*, 14, 149-160.

Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM Review*, 11(2), 226-235.

Wolkowicz, H. (1993). *Measure for symmetric rank-one updates* (Revision of Report CORR 90-03). Waterloo, Ontario: Department of Combinatorics and Optimization, University of Waterloo.