A General Framework for Trust Region Subproblems with Applications to Large Scale Minimization

and
Semidefinite Programming

Henry Wolkowicz
University of Waterloo

(work with Franz Rendl, Univ. of Graz)

The Trust Region Subproblem:

Let

$$q(x) := x^t Ax - 2a^t x,$$

$$(TRS) \qquad \mu^* := \begin{array}{ll} \min & q(x) \\ \text{s.t.} & x^t x = s^2 \quad (\leq s^2). \end{array}$$

where

$A = A^t$, not necessarily semidefinite
$a \in \Re^n$, $s > 0$.
(application: quadratic model for unconstrained minimization)

Main Result: an algorithm that solves TRS using only matrix-vector multiplication; therefore exploits sparsity

Semidefinite Programming Framework

$$
\begin{array}{lll}
 & \max & \operatorname{trace} CZ \\
(SDP) & \text{s.t.} & L(Z) = B \\
 & & Z \succeq 0
\end{array}
$$

$Z \succeq 0$ denotes positive semidefinite
L is a linear operator

Applications: e.g. combinatorial optimization

$$\begin{array}{ll} \max & x^t Q x = \operatorname{trace} Q x x^t \\ \text{s.t.} & x_i = \pm 1,\ x_i^2 = 1\ \forall i \end{array}$$

equivalently (with $X \equiv x x^t$)

$$\begin{array}{lll} & \max & \operatorname{trace} QX \\ (SDP) & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X \succeq 0 \end{array}$$

rank(X)=1 (missing constraint - relaxation)

Other Applications:
Control theory (Lyapunov stability, Ricatti equations), structural computational complexity theory, min-max eigenvalue problems, graph theory

Back to TRS:

$$
(TRS) \qquad \mu^* := \begin{array}{ll} \min & q(x) \\ \text{s.t.} & x^t x = s^2 \quad (\leq s^2). \end{array}
$$

nonlinear least squares, Levenberg 1944, Marquadt 1963

Applications to general minimization, Goldfeld, Quandt and Trotter 1966

Early theoretical results, Forsythe and Golub 1965

Hebden 1973, Reinsch 1967

efficient numerics, Gay 1981

state of the art algorithm, More-Sorenson 1983

Currently:

strong Lagrangian Duality, Stern and Wolkowicz 1993

polynomial time, Ye 1992

at most one local-nonglobal, Martinez 1993

two trust regions, CDT 1984

multiple trust regions: Shor, Ramana, Wolkowicz

Lanczos-eigenvalue approach:
Sorensen 1994,
(with SDP) Rendl and Wolkowicz 1994,
Santos and Sorensen 1995

$$L(x,\lambda) = q(x) - \lambda(x^t x - s^2)$$

Theorem(Stern and Wolkowicz, 1993)
(i) Strong duality holds for TRS, i.e.

$$\mu^* = \min_x \max_\lambda L(x,\lambda) = \max_\lambda \min_x L(x,\lambda).$$

Moreover, attainment holds for x and uniquely for λ.
(ii) A dual problem for TRS, without a duality gap, is

$$(D)\ \max_{A-\lambda I \succeq 0} h(\lambda),$$

where

$$h(\lambda) = \lambda s^2 - a^t (A - \lambda I)^\dagger a, \ \text{(concave)}$$

and $\cdot^\dagger$ denotes the Moore-Penrose generalized inverse.

Proof

$$\begin{aligned}
\mu^* &= \min_{||x||=s} q(x) \\
&\geq \max_{\lambda} \min_{x} L(x,\lambda) \\
&= \max_{A-\lambda I \succeq 0} \min_{x} L(x,\lambda),
\end{aligned}$$

hidden constraints - Hessian $\succeq 0$ and stationarity

global minimum of the Lagrangian is

$$x_\lambda := (A-\lambda I)^\dagger a,$$

Therefore, the right hand side of the above is equal to

$$\begin{aligned}
&= \max_{A-\lambda I \succeq 0} \lambda s^2 - a^t (A-\lambda I)^\dagger a \\
&= \max_{A-\lambda I \succeq 0} h(\lambda),
\end{aligned}$$

This functional is strictly concave and coercive.
$h' = 0$ feasibility

Example (hard case) Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \; a = (1 \; 0)^t, \; s = 1.$$

$A - \lambda I \succeq 0$, implies $\lambda \leq -1$

$\lambda^* = -1$, $\mu^* = -1.5 = h(\lambda^*)$

$x_\lambda = (A - \lambda I)^\dagger a = (.5 \; 0)^t$

$q(x_\lambda) = -.75 > \mu^*$

$||x_\lambda|| < 1$

complementary slackness fails

zero duality gap but no primal attainment

Nonlinear Primal-Dual Pair

The dual problem is

$$\max_{A-\lambda I \succeq 0} h(\lambda) \qquad (D)$$

where

$$h(\lambda) = \lambda s^2 - a^t (A - \lambda I)^\dagger a$$

dual of the dual:

$$\begin{array}{lcl} \min & h(\lambda) + \operatorname{trace} Z(A - \lambda I) & \\ \text{s.t.} & s^2 - ||(A - \lambda I)^\dagger a||^2 - \operatorname{trace} Z = 0 & (DD) \\ & Z \succeq 0. & \end{array}$$

For feasible pair Z, λ, the duality gap is:

$$\operatorname{trace} Z(A - \lambda I)$$

complementary slackness

Gale and More-Sorenson Algorithm explained by SDP Framework

maintain 2nd order cond. $(A - \lambda I) \succ 0$
maximize the dual function $h(\lambda)$
(Newton's method)

solve $0 = h'(\lambda) = s^2 - ||x_\lambda||^2$
better solve $0 = \frac{1}{s} - \frac{1}{||x_\lambda||}$
(almost linear, convex)

Cholesky factoriz: $R^t R = A - \lambda I$
(safeguard positive def.)

Solve	: $R^t R p = a$
Solve	: $R^t q = p$
iterate	: $\lambda \leftarrow \lambda + \left(\frac{\lVert p\rVert}{\lVert q\rVert}\right)^2 \left(\frac{\lVert p\rVert - s}{s}\right)$

Hard Case

indicated by $0 < s^2 - ||x_\lambda||^2$ inside TR

DD feasible

$$s^2 - ||(A - \lambda I)^\dagger a||^2 - \mathrm{trace}\, Z = 0$$

so take a primal step to boundary and try to minimize the duality gap with

$$Z = zz^t, \ \mathrm{trace}\, Z = ||z||^2$$

$$\begin{array}{rcl} q(x + z) & = & (\lambda s^2 - x_\lambda^t (A - \lambda I) x_\lambda) + z^t (A - \lambda I) z \\ & = & h(\lambda) + z^t (A - \lambda I) z. \end{array}$$

$$\min \mathrm{trace}\, Z(A - \lambda I) = z^t (A - \lambda I) z$$

$$A - \lambda I = RR^t \text{ Cholesky decomp}$$

(singular value/vector of R)

Another approach:

Homogenization of TRS

$$\begin{aligned}
\mu^* &= \min_{||x||=s,\ y_0^2=1} x^tAx - 2y_0a^tx \\
&= \max_t \min_{||x||=s} x^tAx - 2y_0a^tx + ty_0^2 - t \\
&= \max_t \min_{||x||=s,\ y_0^2=1} x^tAx - 2y_0a^tx + ty_0^2 - t \\
&= \max_t \min_{||x||^2+y_0^2=s^2+1} x^tAx - 2y_0a^tx + ty_0^2 - t
\end{aligned}$$

$$= \max_t (s^2+1)\lambda_1(D(t)) - t$$

$$D(t) = \begin{pmatrix} t & -a^t \\ -a & A \end{pmatrix}.$$

unconstrained dual problem to TRS

$$D(t) = \begin{pmatrix} t & -a^t \\ -a & A \end{pmatrix}$$

$y = \begin{pmatrix} y_0 \\ x \end{pmatrix}$ normalized eigenvector for $\lambda_{\text{min}} D(t)$

$$k(t) = (s^2 + 1)\lambda_{\text{min}}(D(t)) - t,$$

$$*** \qquad \max_t k(t)$$

Note

$$k'(t) = (s^2 + 1)y_0^2 - 1 = 0$$

is feasibility again for x

SDP Primal-Dual Pair

$$\max_t k(t) = (s^2 + 1)\lambda_{\min}(D(t)) - t,$$

add the variable λ

$$\begin{array}{lc} \max & (s^2 + 1)\lambda - t \\ \text{s.t.} & D(t) \succeq \lambda I \end{array} \quad (DSDP)$$

Lagrangian dual of this dual is:

$$\begin{array}{lc} \min & \operatorname{trace} D(0)X \\ \text{s.t.} & \operatorname{trace} X = s^2 + 1 \\ & X_{11} = 1 \\ & X \succeq 0 \end{array} \quad (PSDP)$$

primal-dual interior point method:

approx. solve perturbed optimality conditions using Newton's method:

$$
\begin{gathered}
\operatorname{trace} X = s^2 + 1 \\
X_{11} = 1 \\
D(t) - \lambda I - Z = 0 \\
\mu Z^{-1} - X = 0 \\
X \succ 0, Z \succ 0
\end{gathered}
$$

Dual Simplex Method

basic feasible dual solution at given t:

$$D(t) - \lambda_{\text{min}}(D(t))I \succeq 0 \text{ and singular}$$

eigenvector $y = \begin{pmatrix} y_0 \\ x \end{pmatrix}$

normalize $y_0 = 1$

test primal feasibility of $X = yy^t$ to get better value for t (interpolation problem)

Alternatively, normalize $||y|| = 1$ with $y_0 \geq 0$, and exploit almost linear structure of

$$\frac{1}{y_0(t)} = \sqrt{s^2 + 1}$$

$$Z = D(t) - \lambda I$$

complementary slackness

$$\operatorname{trace} ZX = 0$$

$$0 = ZX = \begin{bmatrix} t - \lambda & -a^t \\ -a & A - \lambda I \end{bmatrix} \begin{bmatrix} X_{11} & y^t \\ y & \bar{X} \end{bmatrix}.$$

$$t = a^t y + \lambda;$$

$$\bar{X} a = (t - \lambda) y;$$

$$(A - \lambda I) y = a;$$

$$(A - \lambda I) \bar{X} = a y^t;$$

$$A - \lambda I \succeq 0.$$

$$\lambda_i(A) = \alpha_i$$

$$J := \{i : a_i \neq 0\}$$

$$\begin{aligned}
&\det(D(t) - \lambda I) = \\
&\quad = \ (t - \lambda) \prod_{k=1}^{n} (\alpha_k - \lambda) - \sum_{k=1}^{n} a_k^2 \prod_{j \neq k}^{n} (\alpha_j - \lambda) \\
&\quad = \ [t - \lambda - \sum_{j \in J} \frac{a_j^2}{\alpha_j - \lambda}] \prod_{k=1}^{n} (\alpha_k - \lambda) \\
&\quad = \ d(\lambda) \prod_{k=1}^{n} (\alpha_k - \lambda)
\end{aligned}$$

Easy case:

$$\lambda_1(D(t)) \text{ is simple and } \lambda_1(D(t)) < \alpha_1$$

if $\lambda < \alpha_1$ given. Then

$$D(d(\lambda)) - \lambda I \succeq 0 \text{ and singular.}$$

Let $y(t)$ be a normalized eigenvector corresponding to $\lambda_1(D(t))$ and denote its first component by $y_0(t)$. Then $y_0(t) \neq 0$. (wlog > 0)

$y_0(t) : \Re \mapsto (0, 1)$ is strictly monotonically decreasing.

Theorem
$y(t) = \begin{pmatrix} y_0(t) \\ z(t) \end{pmatrix}$ normal eigenvector for

$$\lambda_1(D(t))$$

Then
$y_0(t) \neq 0$ and $v := \frac{1}{y_0(t)} z(t)$ is unique opt of

$$\min\{v^t A v - 2a^t v : v^t v = \frac{1 - y_0(t)^2}{y_0(t)^2}\}.$$

Conversely, if $v \in \Re^n \; and \; \lambda \in \Re$ satisfy

$$(A - \lambda I)v = a; \quad A - \lambda I \succeq 0; \;\; v^t v = \frac{1 - y_0^2}{y_0^2}$$

thereby defining $y_0 > 0$,
then
$y := y_0 \begin{pmatrix} 1 \\ v \end{pmatrix}$
is eigenvector for $D(t)$ for $t := a^t v + \lambda$
and

$$\lambda_1(D(t)) = \lambda$$

Dual-Simplex Method

Initialization:

Find bounds for intervals of uncertainty.

$t_0^l \leq t^* \leq t_0^u$

$\mu^l \leq \mu^* \leq \mu^u$

Begin the main iterations:

while (feasibility or duality gap tols too large)

  Part 1: Find a new estimate t_+.

  $t_+ = \frac{t_c^l + t_c^u}{2}$ (default value)

    Find intersection point of 2 tangent lines

      to the graph of $k(t)$

    update the upper bound μ^u

  if last iterate, on infeasible side, good side

    use inverse interpolation

  elseif on feasible side, bad side

    use inverse interpolation

  endif

  Part 2: Update info for new estimate t_+.

  if $\lambda > 0$ (wrong sign for Lagrange multiplier)

    if $t_+ <= t^*$ then **STOP**

else (correct sign for Lagrange multiplier)
Update lower bound using dual function
endif
if $t_+ < t_*$ (bad side)
Perform a PSDP (negative curvature)
step, update bounds
else
steepest descent step - update bounds
endif
Update the tolerances
endwhile

Tests on SUN SPARC station 1 using MAT-LAB

1. dimensions 1540 to 1565
   30 problems for each dimension
   density of nonzeros .01
   average iterations 4.4453
   average cpu time 54.1876 sec
   average work 1.73 times one Lanczos step

2. dimensions 1540 to 1558
   30 problems for each dimension
   density of nonzeros .013
   average iterations 4.1148
   average cpu time 56.1322 sec
   average work 1.75 times one Lanczos step
   comment: a multiple of the identity was added to get a positive definite Hessian

3. dimensions 1500 to 1565
   2 problems for each dimension
   density of nonzeros .01
   average iterations 4.2951
   average cpu time 57.0429 sec
   average work 4.8589 times one Lanczos step
   comment: hard case; multiplicity of smallest eigenvalue varied from 1 to 6

4. dimensions 1800 to 1865
   12 problems for each dimension
   density of nonzeros .005
   average iterations 4.2453
   average cpu time 28.1692 sec
   average work 4.2589 times one Lanczos step
   comment: hard case; multiplicity of smallest eigenvalue varied from 1 to 6