test

INDEFINITE TRUST REGION SUBPROBLEMS, INTERIOR POINT METHODS, AND MAX-MIN EIGENVALUE PROBLEMS

Henry Wolkowicz
Department of Combinatorics and Optimization
University of Waterloo

work with Ron Stern (Concordia University), Svata Poljak (Charles University), Franz Rendl (University of Graz), Bob Vanderbei (Princeton University)

OUTLINE

- extend trust region theory in two ways: indefinite and 2-sided constraints

- applications to bounds for 0-1 quadratic programming

- interior point numerical algorithm over the positive semidefinite cone of matrices

TRUST REGION METHODS

$$\min_{x \in \Re^n} f(x)$$

Newton's Method (possible indefinite Hessians) model based (at current estimate x_c)

$$f(x) \approx$$

$$m_c(x) := f(x_c) + g_c^t(x - x_c) + \frac{1}{2}(x - x_c)^t B_c (x - x_c)$$

$$B_c \approx \nabla^2 f(x_c), \quad g_c = \nabla f(x_c)$$

TRUST REGION SUBPROBLEM

$$\min q_c(d) := g_c^t d + \frac{1}{2} d^t B_c d$$

$$\text{subject to } ||D_c d|| \leq \delta_c$$

(unlike Newton's method - not scale invariant)

Special Case

$$\mu = \min_{||x||=1} x^t B x$$

$$L(x,\lambda) = x^t B x - \lambda(x^t x - 1) \quad \text{(Lagrangian)}$$

$$\nabla_x L(x,\lambda) = 0 = Bx - \lambda x$$

$$\Longrightarrow Bx = \lambda x \quad \text{(eigenvalue-eigenvector)}$$

$$\mu = x^t B x = \lambda x^t x = \lambda \rightarrow \min$$

We get the smallest eigenvalue.

Nonconvex problem

IFF characterization of **GLOBAL** minimum

Local Minima? (NONE)

General Trust Region Subproblem:

$$
(P) \qquad \begin{array}{lrl} \mu^* := \min & \mu(y) := & y^tBy - 2\psi^t y \\ \text{subject to} & \beta \le & y^tCy \le \alpha, \; y \in \Re^n \end{array}
$$

B, C symmetric, possibly indefinite

$$
-\infty \le \beta \le \alpha \le \infty
$$

- **characterize** global min

- concave max dual problem with no duality gap

- (can characterize local min)
  (connections to eigenvalue perturbation problems)

THEOREM y feasible point and CQ:

$$Cy = 0 \text{ implies } \beta < 0 < \alpha.$$

Then y global minimum
iff
there exists $\lambda \in \Re$ such that

$$(B - \lambda C)y = \psi \text{ (stationarity)},$$

$$B - \lambda C \succeq 0 \text{ (Hessian nonneg def)},$$

and

$$\lambda(\beta - y^t C y) \geq 0 \geq \lambda(y^t C y - \alpha) \quad \begin{pmatrix} \text{compl slack} \\ \text{mult sign} \end{pmatrix}$$

$$B - \lambda C \succ 0 \text{ implies unique min}$$

necessary condition for optimality:

$$\exists\ \hat{\lambda} \in \Re \ \text{ s.t. } \ B - \hat{\lambda} C \geq 0.$$

Define:

regular case or the *positive definite pencil case*

$$\exists\ \hat{\lambda} \in \Re \text{ s.t. } B - \hat{\lambda} C > 0$$

Ignore the irregular case since:

- (i) The set of t where $B - tC$ is positive definite is an open interval which is bounded if and only if C is indefinite.

- (ii) In the irregular case with the function $\det(B - tC)$ not identically 0 in t, there is only one value $\hat{\lambda}$ such that $B - \hat{\lambda} C \geq 0$ holds.

Existence or boundedness
equivalent to
positive semidefinite pencil

THEOREM (EXISTENCE-BOUNDEDNESS)
C nonsingular

- (P) possesses a minimizing point and $\max\{|\alpha|,|\beta|\} > 0$ implies positive semidefinite pencil condition holds.

- Conversely, positive semidefinite pencil condition and regular implies minimizing point exists

DUALITY

We now present a **dual problem** for (P) which is a true **concave maximization** programming problem. This illustrates that (P) is an **implicit convex program** and shows why the **global minimum** can be **characterized** and found efficiently.

Note that if y^* solves (P) with optimal Lagrange multiplier λ^*, then being in the **easy case** implies that $y^* = (B - \lambda^* C)^{-1}\psi$.

THEOREM (DUALITY)

- y^*, λ^*, $\mu^* = \mu(y^*)$ optimal

- Lagrangian function

  $L(y, \nu, \omega) = \mu(y) + \nu(\alpha - y^t C y) + \omega(y^t C y - \beta)$

- Lagrange dual functional

$$\phi(\nu, \omega) = \inf L(y, \nu, \omega)$$

- quadratic dual functional

  $h(\nu, \omega) = \nu\alpha - \omega\beta - \psi^t (B - \nu C + \omega C)^{-1} \psi$

Then

$$\begin{aligned} \mu^* &= \max_{\nu \leq 0, \omega \leq 0} \phi(\nu, \omega) \\ &= \sup_{\substack{B - \nu C + \omega C > 0 \\ \nu \leq 0, \omega \leq 0}} h(\nu, \omega) \end{aligned}$$

Proof If $B - \nu C + \omega C > 0$, then $\phi(\nu, \omega)$ is finite. Moreover, $L(y, \nu, \omega) = \phi(\nu, \omega)$ for $y = (B - \nu C + \omega C)^{-1}\psi$. Substituting for y in L yields

$$h(\nu, \omega) = \phi(\nu, \omega).$$

Now if z is feasible for (P), then for all nonpositive ν, ω we have $\phi(\nu, \omega) \leq L(z, \nu, \omega) \leq \mu(z)$. We now have

$$\begin{aligned} \mu^* &= \min_{z \ feasible} \mu(z) \\ &\geq \sup_{\nu \leq 0, \omega \leq 0} \phi(\nu, \omega), \\ &= \sup_{\substack{B - \nu C + \omega C \geq 0 \\ \nu \leq 0, \omega \leq 0}} \phi(\nu, \omega) \\ &\geq \sup_{\substack{B - \nu C + \omega C > 0 \\ \nu \leq 0, \omega \leq 0}} h(\nu, \omega). \end{aligned}$$

Now, from the optimality conditions, there exists an optimal multiplier λ^*. Let ν^* and ω^* be chosen as in the statement of the Theorem. Then

$$\begin{array}{rcl} \mu^* & = & L(y^*, \nu^*, \omega^*) \\ & = & \phi(\nu^*, \omega^*) \\ & \leq & \max_{\nu \leq 0, \omega \leq 0} \phi(\nu, \omega), \end{array}$$

i.e. this and the previous equations imply that the first equality holds.

If the easy case holds, i.e. $y^* = (B - \lambda^* C)^{-1} \psi$, then

$$h(\nu^*, \omega^*) = L(y^*, \nu^*, \omega^*) = \mu^*.$$

The conclusion follows.

Now suppose that the hard case holds, i.e. $B - \lambda^* C$ is singular. Equivalently, $D - \lambda^* S$ is singular, where $T^t B T = D$ and $T^t C T = S$ are both diagonal, T nonsingular. Let $\bar{y} = T(D - \lambda^* S)^\dagger T^t \psi$, where $\dagger$ denotes the Moore-Penrose generalized inverse. From the optimality conditions, we have that $\psi \in R(B - \lambda^* C)$. Let $\lambda_k \to \lambda^*$ with $B - \lambda_k C$ positive definite. Let ν_k, ω_k correspond to λ_k as ν^*, ω^* corresponds to λ^*. Then, from the simultaneous diagonalization, we conclude that

$$y_k = (B - \lambda_k C)^{-1} \psi = T(D - \lambda_k S)^{-1} T^t \psi \to \bar{y}.$$

Moreover, $y^* = \bar{y} + z$ for some z in the null space of $B - \lambda^* C$, and we also have $z \perp \psi$. Now $L(y, \nu, \omega) = y^t(B - \nu C - \omega C)y - 2\psi^t y + \nu\alpha - \omega\beta$. So

$$h(\nu_k, \omega_k) = L(y_k, \nu_k, \omega_k) \to L(\bar{y}, \nu^*, \omega^*)$$

and this equals

$$L(y^*, \nu^*, \omega^*) = \mu^*.$$

Attainment follows directly from the optimality conditions.

$\square$

duality result with only one multiplier

COROLLARY

$$\bar{h}(\lambda) = -(-\lambda)_+\alpha + (\lambda)_+\beta - \psi^t(B - \lambda C)^\dagger\psi,$$

where

$$(\lambda)_+ = \begin{cases} \lambda & \text{if } \lambda \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mu^* = \max_{B-\lambda C \succ 0} \bar{h}(\lambda).$$

Standard trust region subproblem

COROLLARY $C = I$, $\beta < 0 < \alpha$

$$\bar{g}(\lambda) = \lambda\alpha - \psi^t (B - \lambda C)^\dagger \psi.$$

Then

$$\mu^* = \sup_{B - \lambda C \succ 0,\ \lambda \le 0} \bar{g}(\lambda).$$

In addition, in the hard case,

$$\mu^* = \sup_{B - \lambda C \succ 0,\ \lambda \le 0} \bar{g}(\lambda) = \max_{B - \lambda C \succeq 0,\ \lambda \le 0} \bar{g}(\lambda).$$

LOCAL-NONGLOBAL MINIMA
(extension of results of Mario Martinez)

LEMMA Suppose that y^* is a local-nonglobal minimizer of (P) and suppose that the pencil $B - \lambda C$ is regular and that the constraint qualification

$$Cy^* \neq 0$$

holds. Then

$$\lambda_{n-1}(B - \lambda^* C) > 0 > \lambda_n(B - \lambda^* C).$$

There are at most two local-nonglobal minima.
The existence of lngm implies easy case.
A local duality theory exists.

± 1 QUADRATIC PROGRAMMING

- $$(P) \quad \mu^* := \max \; q(x) := x^t Q x + c^t x, \\ x \in F := \{-1, 1\}^n,$$

  Q $n \times n$ symmetric matrix, $c \in \Re^n$.

- Relaxed Problem - $K \supset F$

  $$(RP) \quad f(u) = \max_{x \in K} q_u(x) := \\ x^t(Q - \mathrm{diag}(u))x + u^t e + c^t x,$$

- Solve Tractable Problem

  $$B := \min_{u \in L} f(u)$$

- Upper Bound

  $$\mu^* \leq B$$

Three Relaxations

---

relaxed problem

$$(RP_v^1) \quad f_1(v) := \max_{-1 \leq x \leq 1} q_v(x).$$

bound

$$B_1 := \min_{Q - \text{diag}(v) \leq 0} f_1(v).$$

---

relaxed problem

$$(RP_u^2) \quad f_2(u) := \max_{||y||^2 = n} q_u(y).$$

bound

$$B_2 := \min_{u^t e = 0} f_2(u) = \min_u f_2(u).$$

$$Q^c := \begin{bmatrix} 0 & \frac{1}{2}c^t \\ \frac{1}{2}c & Q \end{bmatrix}.$$

$$q_u^c(y) := y^t(Q^c - \mathrm{diag}(u))y + u^t e$$

---

relaxed problem

$$(RP_u^3) \qquad f_3(u) := \max_{||y||^2=n+1} q_u^c(y)$$
$$= (n+1)\lambda_{\max}(Q^c - \mathrm{diag}(u)) + u^t e.$$

bound

$$B_3 := \min_{u^t e = 0} f_3(u) = \min_u f_3(u).$$

BOUND 1 - Convex Quadratic Programming

Consider the shifted function

$$q_v(x) := x^t(Q - \mathrm{diag}(v))x + v^t e + c^t x,$$

and the relaxed problem

$$(RP_v^1) \quad f_1(v) := \max_{-1 \le x \le 1} q_v(x).$$

Then a bound for (P) is

$$B_1 := \min_{Q - \mathrm{diag}(v) \le 0} f_1(v).$$

Properties for Bound B_1.

1. $\mu^* \leq B_1$.

2. f_1 is convex, finite valued, with subdifferential

$$\partial f_1(v) = \text{conv}\{z = (1 - x_i^2) \in \Re^n : \\ x = (x_i) \in \Re^n - 1 \leq x \leq 1,\ f_1(v) = q_v(x)\}.$$

3. B_1 is attained for some $v \in \Re^n$ such that $\lambda_{max}(Q - \text{diag}(v)) = 0$.

4.

$$B_1 = \inf_{Q - \text{diag}(v) < 0} f_1(v).$$

Lagrangian

$$L_1(v,\Lambda) := f_1(v) + \mathrm{trace}\Lambda(Q - \mathrm{diag}(v)),$$

where Λ is a symmetric, positive semidefinite Lagrange multiplier matrix.

Optimality conditions:

$$\begin{array}{cc} 0 \in \partial f_1(v) - \mathrm{diag}(\Lambda) & \text{(stationarity)} \\ \mathrm{trace}\Lambda(Q - \mathrm{diag}(v)) = 0 & \text{(complementary slack)} \\ \Lambda = \Lambda^t \succeq 0 & \text{(multiplier sign).} \end{array}$$

BOUND 2 - Optimization Over Sphere

$$q_u(y) := y^t(Q - \mathrm{diag}(u))y + u^t e + c^t y,$$

the second relaxed problem

$$(RP_u^2) \quad f_2(u) := \max_{||y||^2 = n} q_u(y).$$

Now a bound for (P) is

$$B_2 := \min_{u^t e = 0} f_2(u) = \min_u f_2(u).$$

Properties for bound B_2.

1. $\mu^* \leq B_2$.

2. f_2 is convex, finite valued, with subdifferential

$$\partial f_2(u) = \text{conv}\{z = (1 - y_i^2) \in \Re^n : \\ y = (y_i) \in \Re^n,\ ||y||^2 = n,\ f_2(u) = q_v(y)\}.$$

3. The bound B_2 is attained for some $u \in \Re^n$. Moreover, if $B_2 > \mu^*$, then the hard case holds for (RP_u^2).

THEOREM

The bound $B_1 = B_2$.

In fact:

$$\begin{array}{rcl} f_2(u) & = & q_u(y_u) \\ & \geq & \max_{(-1 \leq y \leq 1)} q_{(u+\lambda e)}(y) \\ & = & f_1(u+\lambda e) \end{array}$$

and

$B_2 = f_2(u)$, with Lagrange multiplier λ for (RP_u^2) if and only if $B_2 = f_2(u) = f_1(u+\lambda e) = B_1$.

THEOREM

Suppose that $f_4(v) := \max_{-1 \le x \le 1} q_v^c(x)$ and

$$B_4 := \min_{Q^c - \mathrm{diag}(v) \le 0} f_4(v).$$

Then

$$B_1 = B_2 = B_3 = B_4.$$

The *max-min eigenvalue problem* is:

$$(MMP) \quad \omega^* := \max_{v^t e = 0} \lambda_{\min}(C - \mathrm{Diag}(v)).$$

It can be formulated as:

$$\omega^* = \max\{\omega : v^t e = 0,\ C - \mathrm{Diag}(v) \succeq \omega I\}.$$

With $y = v + \omega e$, an equivalent problem is:

$$\max\{e^t y :\ C - \mathrm{Diag}(y) \succeq 0\}.$$

The dual is the min-max of the Lagrangian

$$\min_{X^a \succeq 0} \max_{y}\ e^t y + \mathrm{trace} X(C - \mathrm{Diag}(y)).$$

or:

$$\begin{array}{ll} \min & \mathrm{trace} CX \\ \text{subject to} & \mathrm{diag}(X) = e \\ & X \succeq 0. \end{array}$$

We get the dual pair of "linear type programs"

$$\begin{array}{ll} \max & e^t y \\ \text{subject to} & C - \mathrm{Diag}(y) - Z = 0 \\ & Z \succeq 0. \end{array}$$

$$\begin{array}{ll} \min & \mathrm{trace}\, CX \\ \text{subject to} & \mathrm{diag}(X) = e \\ & X \succeq 0. \end{array}$$

PRIMAL-DUAL INTERIOR POINT METHOD

The primal-dual log-barrier problem is then

$$(B) \qquad \begin{array}{ll} \omega_\mu^* = \max & e^t y + \mu \log\det(Z) \\ \text{subject to} & C - \mathrm{Diag}(y) - Z = 0. \end{array}$$

The Lagrangian $L(y, Z, X)$ is

$$e^t y + \mu \log\det(Z) + \mathrm{trace}X(C - \mathrm{Diag}(y) - Z),$$

while the optimality conditions $F(y, Z, X) = 0$ are:

$$\begin{array}{rcll} e - \mathrm{diag}(X) & = & 0 & \text{(primal feasibility)} \\ -X + \mu Z^{-1} & = & 0 & \text{(complementary slackness)} \\ C - \mathrm{Diag}(y) - Z & = & 0 & \text{(dual feasibility)} \\ Z \succeq 0, \; X \succeq 0. & & & \end{array}$$

Apply Newton's method:

$$\begin{bmatrix} 0 & 0 & \mathsf{diag}(\cdot) \\ 0 & \mu Z^{-1} \cdot Z^{-1} & I \\ \mathsf{Diag}(\cdot) & I & 0 \end{bmatrix} \begin{pmatrix} \delta y \\ \delta Z \\ \delta X \end{pmatrix} =$$

$$\begin{pmatrix} (e - \mathsf{diag}(X)) \\ (-X + \mu Z^{-1}) \\ (C - \mathsf{Diag}(y) - Z) \end{pmatrix}.$$

simplify with $Z = C - \mathsf{Diag}(y)$:

$$\begin{array}{ccc} \delta y & = & (\mu Z^{-1} \circ Z^{-1})^{-1}(e - \mu \mathsf{diag}(Z^{-1})); \\ \delta Z & = & -\mathsf{Diag}(\delta y); \\ X + \delta X & = & \mu Z^{-1}(Z + \mathsf{Diag}(\delta y))Z^{-1}. \end{array}$$