School of Mathematics

Linear Algebra Techniques in Interior Point Methods for Optimization

Jacek Gondzio
Email: J.Gondzio@ed.ac.uk
URL: http://www.maths.ed.ac.uk/~gondzio

NA Conference, Dundee, June 2003

- Linear Algebra in IPMs
- LP, QP, NLP: Linear Algebra is the same
- Symmetric Systems:
  - Positive Definite vs Indefinite Systems
  - Quasi-definite Systems
  - Primal and Dual Regularization
- Unavoidable Ill-conditioning
  - IPM Scaling Matrices
  - Dikin's Bound
- Primal-Dual Regularized Factorization
- Exploiting Structure in IPMs

Linear Algebra of IPM for LP

First order optimality conditions

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \mu e. \end{aligned}$$

Newton's direction

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix},$$

where

$$\begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^T y - s \\ \mu e - XSe \end{bmatrix}.$$

Use the third equation to eliminate

$$\begin{aligned} \Delta s &= X^{-1}(\xi_\mu - S\Delta x) \\ &= -X^{-1} S \Delta x + X^{-1}\xi_\mu, \end{aligned}$$

from the second equation and get

$$\begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1}\xi_\mu \\ \xi_p \end{bmatrix}.$$

where $\Theta = XS^{-1}$ is a diagonal scaling matrix.

IPMs: LP, QP & NLP

Augmented system in **LP**

$$\begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r \\ h \end{bmatrix}.$$

Eliminate Δx from the first equation and get normal equations

$$(A\Theta A^T)\Delta y = g.$$

Augmented system in **QP**

$$\begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r \\ h \end{bmatrix}.$$

Eliminate Δx from the first equation and get normal equations

$$(A(Q + \Theta^{-1})^{-1} A^T)\Delta y = g.$$

Augmented system in **NLP**

$$\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r \\ h \end{bmatrix}.$$

Eliminate Δx from the first equation and get normal equations

$$(AQ^{-1}A^T + \Theta^{-1})\Delta y = g.$$

Two step solution method:

- factorization to LDL^T form,
- backsolve to compute direction Δy.

Two options are possible:

1. Replace diagonal matrix D with a block-diagonal one and allow 2×2 (indefinite) pivots

$$\begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & a \\ a & d \end{bmatrix}.$$

Hence obtain a decomposition $H = LDL^T$ with **block-diagonal** D.

2. Regularize indefinite matrix to produce a **quasidefinite** matrix

$$K = \begin{bmatrix} -E & A^T \\ A & F \end{bmatrix},$$

where
$E \in \mathcal{R}^{n \times n}$ is positive definite,
$F \in \mathcal{R}^{m \times m}$ is positive definite, and
$A \in \mathcal{R}^{m \times n}$ has full row rank.

A symmetric matrix is called **quasidefinite** if

$$K = \begin{bmatrix} -E & A^T \\ A & F \end{bmatrix},$$

where $E \in \mathcal{R}^{n \times n}$ and
$F \in \mathcal{R}^{m \times m}$ are positive definite, and
$A \in \mathcal{R}^{m \times n}$ has full row rank.

Symmetric nonsingular matrix K is *factorizable* if there exists a diagonal matrix D and a unit lower triangular matrix L such that $K = LDL^T$.

The symmetric matrix K is *strongly factorizable* if for any permutation matrix P a factorization $PKP^T = LDL^T$ exists.

Vanderbei (1995) proved that
Symmetric QDFM's are strongly factorizable.
SIOPT 5 (1995) 100-113.

For any quasidefinite matrix
there exists a **Cholesky-like** factorization

$$\bar{H} = LDL^T,$$

where
D is **diagonal** but **not positive definite**:
has n negative pivots;
and m positive pivots.

From Indefinite to Quasidefinite Matrix

Indefinite matrix

$$H = \begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix}.$$

Vanderbei *SIOPT* 5 (1995) 100-113.
Replace $Ax = b$ with $Ax + s = b$

$$H_V = \begin{bmatrix} -\Theta_s^{-1} & 0 & I \\ 0 & -Q - \Theta^{-1} & A^T \\ I & A & 0 \end{bmatrix}$$

and eliminate Θ_s^{-1}

$$K = \begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & \Theta_s \end{bmatrix}.$$

Saunders (1996) SIAM Adams & Nazareth (eds)

$$H_S = \begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} + \begin{bmatrix} -\gamma^2 I_n & 0 \\ 0 & \delta^2 I_m \end{bmatrix},$$

where $\gamma\delta \geq \sqrt{\varepsilon} = 10^{-8}$.

Altman & Gondzio *OMS* 11-12 (99) 275-302.
Use **dynamic regularization**

$$\bar{H} = \begin{bmatrix} -\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} + \begin{bmatrix} -R_p & 0 \\ 0 & R_d \end{bmatrix},$$

$R_p \in \mathcal{R}^{n \times n}$ is a *primal* regularization
$R_d \in \mathcal{R}^{m \times m}$ is a *dual* regularization.

Primal Regularization

Primal barrier problem

$$\begin{array}{ll} \min & z_P = c^T x + \frac{1}{2} x^T Q x - \mu \sum_{j=1}^n (\ln x_j + \ln s_j) \\ \text{s. to} & Ax = b, \\ & x + s = u, \\ & x, s > 0 \end{array}$$

$$\begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ h \end{bmatrix}.$$

Primal **regularized** barrier problem

$$\begin{array}{ll} \min & z_P + \frac{1}{2}(x - x_0)^T R_p (x - x_0) \\ \text{s. to} & Ax = b, \\ & x + s = u, \\ & x, s > 0 \end{array}$$

$$\begin{bmatrix} -Q - \Theta^{-1} - R_p & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f' \\ h \end{bmatrix},$$

where

$$f' = f - R_p(x - x_0).$$

Dual barrier problem

$$\begin{array}{ll} \max & z_D = b^T y - u^T w - \frac{1}{2} x^T Q x + \mu \sum_{j=1}^{n} (\ln z_j + \ln w_j) \\ \text{s. to} & A^T y + z - w - Qx = c, \\ & x \geq 0, z, w > 0 \end{array}$$

$$\begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ h \end{bmatrix}.$$

Dual **regularized** barrier problem

$$\begin{array}{ll} \max & z_D - \frac{1}{2} (y - y_0)^T R_d (y - y_0) \\ \text{s. to} & A^T y + z - w - Qx = c, \\ & x \geq 0, z, w > 0 \end{array}$$

$$\begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & R_d \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ h' \end{bmatrix},$$

where

$$h' = h - R_d (y - y_0).$$

Higher Order Primal Dual Method

Problem	Dimensions				LOQO	HOPDM
	m	n	nz(A)	nz(Q)	nz(L)	nz(L)
nug12	3192	8856	44244	0	3091223	1969957
nug15	6330	22275	110700	0	-	7374972
cvxqp1_m	500	1000	1498	2984	71487	75973
cvxqp1_l	5000	10000	14998	29984	4056820	3725045
cvxqp2_m	250	1000	749	2984	52917	51923
cvxqp2_l	2500	10000	7499	29984	2923584	2754141
cvxqp3_m	750	1000	2247	2984	79957	90433
cvxqp3_l	7500	10000	22497	29984	4411197	4291057

200 MHz Pentium II PC, Linux.

Problem	LOQO		HOPDM	
	iters	time	iters	time
nug12	24	4417.7	13	1140.3
nug15	-	-	15	10276.6
cvxqp1_m	32	13.78	9	6.63
cvxqp1_l	72	18361.1	11	2874.4
cvxqp2_m	16	4.06	9	4.02
cvxqp2_l	25	3849.4	8	1353.7
cvxqp3_m	49	25.45	9	9.11
cvxqp3_l	100	27447.6	8	2461.2

Ill-conditioning

Assume **Normal Equations** are used in LP and the feasible IPM is used ($\xi_p = 0$ and $\xi_d = 0$)

$$(A \Theta A^T) \Delta y \;=\; A \Theta r,$$

where $\Theta = X S^{-1}$ and $r = -X^{-1} \xi_\mu$.

Optimal Partition:

Basic variables	$x_B \to x_B^* > 0$	$s_B \to s_B^* = 0$
Non-basic variables	$x_N \to x_N^* = 0$	$s_N \to s_N^* > 0$

For **basic** variables: $\Theta_j = x_j / s_j \to \infty$;
For **non-basic** variables: $\Theta_j = x_j / s_j \to 0$.

Hence

$$A \Theta A^T = \sum_{j \in \mathcal{B}} \theta_j a_{.j} a_{.j}^T + \sum_{j \in \mathcal{N}} \theta_j a_{.j} a_{.j}^T \to \sum_{j \in \mathcal{B}} \theta_j a_{.j} a_{.j}^T.$$

The matrix $H = A \Theta A^T$ usually has a huge condition number $\kappa(H)$. Although $\kappa(H) \gg 1/\epsilon$, where ϵ is the relative precision of the computer (e.g. $\epsilon = 10^{-16}$), IPMs do converge.

Dikin's Bound

Theorem: (Dikin, 1974)
Upravlaemye Sistemy 12 (1974) pp 54-60.

Let $A \in \mathcal{R}^{m \times n}$ be a full row rank matrix;
g be a vector of dimension n; and
D_+ be the set of $n \times n$ diagonal positive definite matrices.
Then

$$\begin{aligned} \sup_{D \in D_+} \| (A D A^T)^{-1} A D g \| &= \max_{\mathcal{J} \in \mathcal{J}(A)} \| A_{\mathcal{J}}^{-T} g_{\mathcal{J}} \| \\ \sup_{D \in D_+} \| (A D A^T)^{-1} A D \| &= \max_{\mathcal{J} \in \mathcal{J}(A)} \| A_{\mathcal{J}}^{-T} \| \end{aligned}$$

where $\mathcal{J}(A)$ is the set of column indices associated with nonsingular $m \times m$ submatrices of A.

Corollary:
The linear system arising in IPMs for LP

$$(A \Theta A^T) \Delta y \;=\; A \Theta r,$$

produces more accurate solutions than those one could have expected from a "classical" worst-case analysis.

Forsgren and Sporre (2001) generalized Dikin's result for a subclass of positive definite weight matrices W. *SIMAX* 22 (2001) 42-56.

Lemma:
Let $A \in \mathcal{R}^{m \times n}$ be a full row rank matrix;
g be a vector of dimension n; and
W_+ be the set of $n \times n$ matrices defined as

$$W = \sum_{i=1}^{k} \alpha_i W_i,$$

where $\alpha_i > 0$ and $W_i = U_i D_i U_i^T$ with U_i bounded and D_i diagonal positive definite $\forall i = 1, ..., k$.
Then

$$\sup_{W \in W_+} \|(AWA^T)^{-1} AWg\|$$

$$\sup_{W \in W_+} \|(AWA^T)^{-1} AW\|$$

are bounded.

This Lemma extends Dikin's result to quadratic and nonlinear optimization.

The Lemma does not hold for arbitrary positive definite matrix W.

Interior Point Methods:

- are well-suited to large-scale optimization
- can take advantage of the parallelism

Large problems are "structured":

- partial separability
- spatial distribution
- dynamics
- uncertainty
- etc.

Object-Oriented Parallel Solver (OOPS)

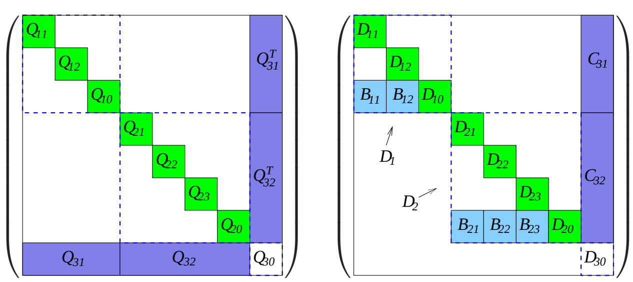
- Exploits structure
- Runs in parallel
- Solves problems with millions of variables

Andreas Grothey will talk about OOPS.

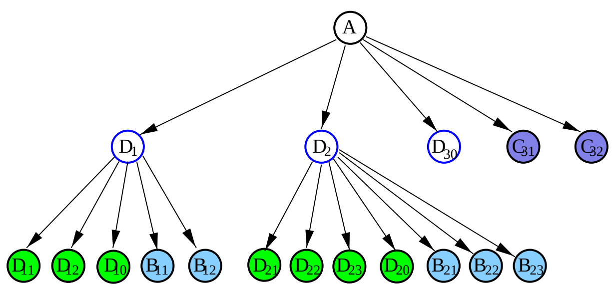
Gondzio & Sarkissian:
Math Prog 96 (2003) 561-584.
Gondzio & Grothey:
SIOPT 13 (2003) 842-864.

Tree Description of Block-Structures

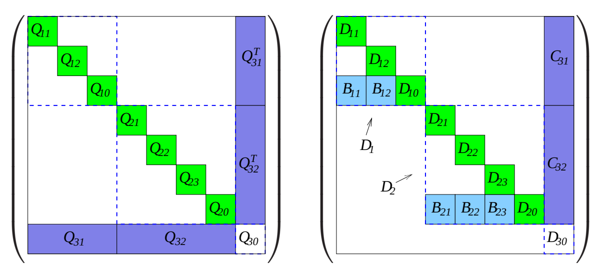
Structured Matrix:

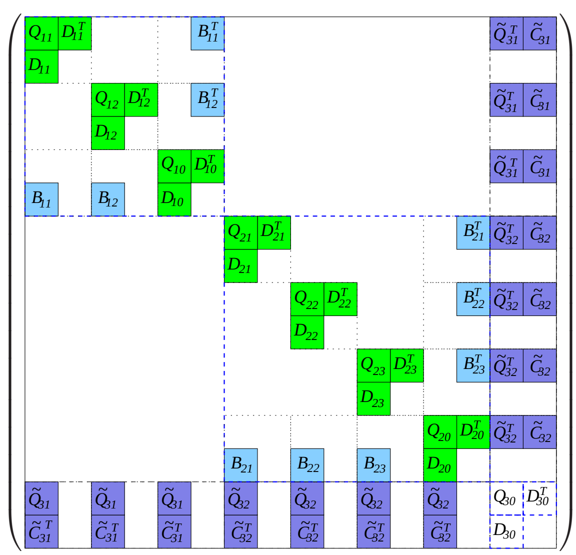


Associated Tree:



Reordered Augmented Matrix





Cholesky factors sometimes get hopelessly dense.

QAP (Quadratic Assignment Problems) and NUG problems (dual QAPs)

Prob	Dimensions		
	rows	columns	nonzeros
qap12	3192	8856	38304
qap15	6330	22275	94950
nug12	3192	8856	38304
nug15	6330	22275	94950

Normal Equations:

Prob	nz(AAt)	nz(LLt)	Flops
qap12	74592	2135388	2.378e+9
qap15	186075	8191638	1.792e+10
nug12	74592	2789960	4.014e+9
nug15	186075	11047639	3.240e+10

Augmented System:

Prob	nz(A)	nz(LLt)	Flops
qap12	38304	1969957	2.046e+9
qap15	94950	7374972	1.522e+10
nug12	38304	1969957	2.046e+9
nug15	94950	7374972	1.522e+10

- Unavoidable Ill-conditioning:
  - benign in direct approach;
  - challenge for iterative approach.
- Positive Definite vs Indefinite Systems
- Preconditioners for Structured Matrices
- Preconditioners for Indefinite System
  - Motivation
    - ∗ Sparsity Issues
    - ∗ Numerical Properties
  - Spectral Analysis
  - Influence of Regularizations
- Conclusions
- What's to Come in IPMs

Iterative Methods

Normal Equations or **Augmented System**:

- NE is positive definite:
  can use conjugate gradients;
- AS is indefinite:
  can use BiCGSTAB, GMRES, QMR;

AS is generally more flexible.
Oliveira (1997) PhD Thesis, Rice Univ.
Oliveira & Sorensen (1997) TR, Rice Univ.
$\longrightarrow$ It is better to precondition AS.

O, OS show that all preconditioners for the NE have an equivalent for the AS while the opposite is not true.

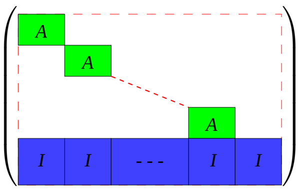
After all, NE is equivalent to a restricted order of pivoting in AS.

Iterative Methods

Many attempts (LP, QP, NLP and PDE):

- Gill, Murray, Ponceleon, Saunders
  SIMAX 13 (1992) 292-311.
- Lukšan & Vlček
  NLAA 5 (1998) 219-247.
- Golub & Wathen
  SISC 19 (1998) 530-539.
- Murphy, Golub & Wathen
  SISC 21 (2000) 1969-1972.
- Keller, Gould & Wathen
  SIMAX 21 (2000) 1300-1317.
- Perugia & Simoncini
  NLAA 7 (2000) 585-616.
- Castro
  SIOPT 10 (2000) 852-877.
- Gould, Hribal & Nocedal
  SISC 23 (2001) 1376-1395.
- Durazzi & Ruggiero
  NLAA (to appear).
- Rozložník & Simoncini
  SIMAX 24 (2002) 368-391.

Castro *SIOPT* 10 (2000) 852-877.



Normal-equations matrix

$$\begin{bmatrix} A_1A_1^T & & & & A_1B_1^T \\ & A_2A_2^T & & & A_2B_2^T \\ & & \ddots & & \vdots \\ & & & A_nA_n^T & A_nB_n^T \\ B_1A_1^T & B_2A_2^T & \cdots & B_nA_n^T & \sum_{i=1}^{n+1} B_iB_i^T \end{bmatrix} = \begin{bmatrix} E & B^T \\ B & F \end{bmatrix},$$

where E and F are positive definite.

E is easily invertible (block-diagonal).
The inverse of Schur complement matrix $F - BE^{-1}B^T$ can be written as the power series:

$$(F - BE^{-1}B^T)^{-1} = \sum_{i=0}^{\infty} (F^{-1}BE^{-1}B^T)^i F^{-1}.$$

Finite approximation of the series:
$\rightarrow$ Very efficient preconditioner.

Murphy, Golub & Wathen
SISC 21 (2000) 1969-1972.

Consider a matrix

$$H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix},$$

where
$Q \in \mathcal{R}^{n \times n}$ is positive definite, and
$A \in \mathcal{R}^{m \times n}$ has full row rank.

Consider the preconditioner which incorporates an exact Schur complement $AQ^{-1}A^T$.
For example:

$$P_1 = \begin{bmatrix} Q & 0 \\ 0 & AQ^{-1}A^T \end{bmatrix} \quad \text{or} \quad P_2 = \begin{bmatrix} Q & A^T \\ 0 & AQ^{-1}A^T \end{bmatrix}.$$

The preconditioned matrices $P^{-1}H$ have only two or three distinct eigenvalues.

MGW conclude:
"The approximations of the Schur complement lead to preconditioners which can be very effective even though they are in no sense approximate inverses".

CG with Indefinite Preconditioner

Consider the indefinite matrix

$$H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix},$$

where
$Q \in \mathcal{R}^{n \times n}$ is positive definite, and
$A \in \mathcal{R}^{m \times n}$ has full row rank.
The CG method may fail when applied to an indefinite system.

Rozložník & Simoncini
SIMAX 24 (2002) 368-391.

RS consider the preconditioner P which guarantees that all eigenvalues of the preconditioned matrix $P^{-1}H$ are positive and bounded away from zero.

Although $P^{-1}H$ is indefinite

- the CG can be applied to this problem,
- the asymptotic rate of convergence of CG is approximately the same as that obtained for a positive definite matrix with the same eigenvalues as the original system.

Indefinite Block Preconditioner

Consider again the matrix

$$H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix},$$

where
$Q \in \mathcal{R}^{n \times n}$ is positive definite, and
$A \in \mathcal{R}^{m \times n}$ has full row rank.

Consider a preconditioner of the form:

$$P = \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix},$$

where $D \in \mathcal{R}^{n \times n}$ is positive definite.

Keller, Gould & Wathen
SIMAX 21 (2000) 1300-1317.

Theorem. *Assume that A has rank m $(m < n)$. Then, $P^{-1}H$ has at least $2m$ unit eigenvalues, and the other eigenvalues are positive and satisfy*

$$\lambda_{min}(D^{-1}Q) \;\le\; \lambda \;\le\; \lambda_{max}(D^{-1}Q).$$

Proof: The preconditioned matrix (left) reads

$$\begin{aligned} P^{-1}H &= \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} = \\ &= \begin{bmatrix} D^{-1} - D^{-1}A^TM^{-1}AD^{-1} & D^{-1}A^TM^{-1} \\ M^{-1}AD^{-1} & -M^{-1} \end{bmatrix} \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \\ &= \begin{bmatrix} D^{-1}Q - D^{-1}A^TM^{-1}AU & 0 \\ M^{-1}AU & I_m \end{bmatrix} = \begin{bmatrix} X & 0 \\ Y & I_m \end{bmatrix}, \end{aligned}$$

where $M = AD^{-1}A^T$, $U = D^{-1}Q - I$.

$P^{-1}H$ has m linearly independent eigenvectors associated with the eigenvalue $\lambda = 1$ since for $w_i \in \mathcal{R}^m$

$$P^{-1}H \begin{bmatrix} 0 \\ w_i \end{bmatrix} = \begin{bmatrix} 0 \\ w_i \end{bmatrix}.$$

The remaining n eigenvectors are the same as those of the matrix $X = D^{-1}Q - D^{-1}A^TM^{-1}AU$.

Matrix X has at least m other unit eigenvalues. Indeed, for any $x \in \mathcal{R}^m$ we write

$$\begin{aligned} X^TA^Tx &= (I + U^T(I - A^TM^{-1}AD^{-1}))A^Tx = \\ &= A^Tx + U^T(A^Tx - A^Tx) = A^Tx. \end{aligned}$$

The remaining $n{-}m$ eigenvalues and eigenvectors of $P^{-1}H$ have to satisfy

$$\begin{aligned} Qx \;+A^Ty &= \lambda Dx \;+\lambda A^Ty \\ Ax &= \lambda Ax. \end{aligned}$$

If $\lambda \neq 1$ the second equation yields $Ax = 0$.
Let us multiply the first equation by x^T.
Recalling that $x^TA^T = 0$ we obtain

$$x^TQx = \lambda x^TDx, \quad \Rightarrow \quad \lambda = \frac{x^TQx}{x^TDx} = q(D^{-1}Q).$$

The last expression is the Rayleigh quotient of the generalized eigenproblem $Dv = \mu Qv$. Since both D and Q are positive definite we have for every $x \in \mathcal{R}^n$

$$0 < \lambda_{\text{min}}(D^{-1}Q) \le \frac{x^TQx}{x^TDx} \le \lambda_{\text{max}}(D^{-1}Q)$$

and finally

$$\lambda_{\text{min}}(D^{-1}Q) \le \lambda \le \lambda_{\text{max}}(D^{-1}Q).$$

Conclusion:
The preconditioner satisfies the requirements of **Rozložník & Simoncini**.

How to choose D?

Bergamaschi, Gondzio & Zilli,
Preconditioning indefinite systems in interior point methods for optimization,
Tech. Rep. MS-02-02.

Augmented system in **QP, NLP**

$$H = \begin{bmatrix} -Q - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix}.$$

Drop off-diagonal elements from Q:
Replace

$$-Q - \Theta^{-1}$$

with

$$D = -diag(Q) - \Theta^{-1}.$$

Preconditioners: Motivation

Sparsity issues: irreducible blocks in QP.

Consider the matrices

$$Q = \begin{bmatrix} \mathbf{x} & \mathbf{x} & & & \\ \mathbf{x} & \mathbf{x} & & & \\ & & x & & \\ & & & x & \\ & & & & x \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} x & & & & x \\ x & & x & & \\ & x & x & & x \\ & x & & x & \end{bmatrix},$$

giving

$$H = \begin{bmatrix} \mathbf{x} & \mathbf{x} & & & & x & x & & \\ \mathbf{x} & \mathbf{x} & & & & & & x & x \\ & & x & & & & x & x & \\ & & & x & & & & & x \\ & & & & x & x & & x & \\ x & & & & x & & & & \\ x & & x & & & & & & \\ & x & x & & x & & & & \\ & x & & x & & & & & \end{bmatrix}.$$

If the elimination starts from h_{11} or h_{22}, then

$$H = \begin{bmatrix} \mathbf{x} & \mathbf{x} & & & & x & x & \mathrm{f} & \mathrm{f} \\ \mathbf{x} & \mathbf{x} & & & & \mathrm{f} & \mathrm{f} & x & x \\ & & x & & & & x & x & \\ & & & x & & & & & x \\ & & & & x & x & & x & \\ x & \mathrm{f} & & & x & & & & \\ x & \mathrm{f} & x & & & & & & \\ \mathrm{f} & x & x & & x & & & & \\ \mathrm{f} & x & & x & & & & & \end{bmatrix}.$$

Conclusion:
Drop off-diagonal elements form Q.

D is a diagonal matrix
$\longrightarrow$ Free choice between NE and AS.

Preconditioner 1
Compute the Cholesky-like factorization.

$$P_1 = \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix} = L\bar{D}L^T.$$

Preconditioner 2
Reduce the system to Normal Equations $AD^{-1}A^T$, compute the Cholesky factorization

$$AD^{-1}A^T = L_0 D_0 L_0^T,$$

and use:

$$P_2 = \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ AD^{-1} & L_0 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & -D_0 \end{bmatrix} \begin{bmatrix} I & D^{-1}A^T \\ 0 & L_0^T \end{bmatrix}.$$

The **regularization**

$$\bar{H}_R = \begin{bmatrix} -Q & A^T \\ A & 0 \end{bmatrix} + \begin{bmatrix} -R_p & 0 \\ 0 & R_d \end{bmatrix},$$

changes the **eigenvalues** of the preconditioned matrix:

without the regularization:

$$\lambda(P^{-1}H) = \frac{x^TQx}{x^TDx}$$

with the regularization:

$$\lambda(P_R^{-1}H_R) = \frac{-x^TQx + \delta}{-x^TDx + \delta},$$

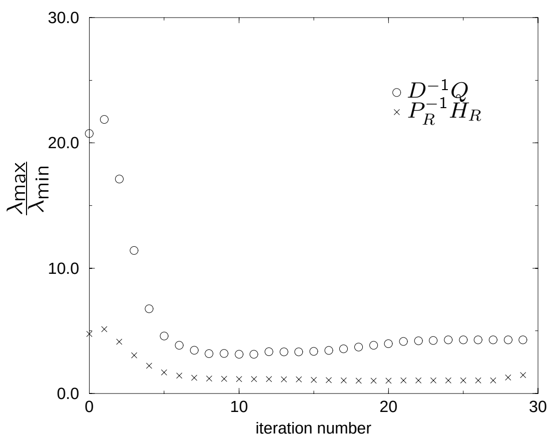
where $\delta = x^TR_p x + y^TR_d y > 0$.

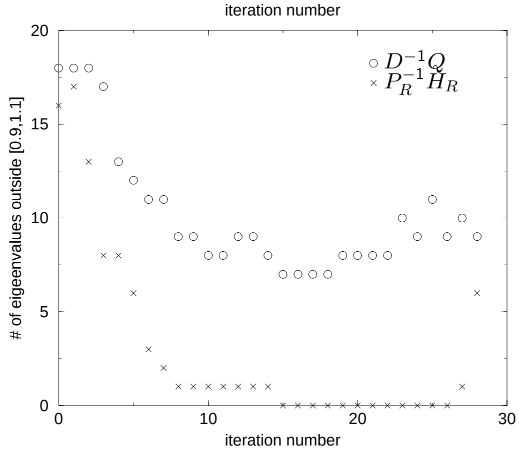
For any $\alpha, \beta, t > 0$, the function $h(t) = \frac{\alpha+t}{\beta+t}$ is *increasing* if $\frac{\alpha}{\beta} \leq 1$, and *decreasing* if $\frac{\alpha}{\beta} > 1$.

Hence:
if $\lambda(P^{-1}H) < 1$, then $\lambda(P_R^{-1}H_R) > \lambda(P^{-1}H)$.
if $\lambda(P^{-1}H) > 1$, then $\lambda(P_R^{-1}H_R) < \lambda(P^{-1}H)$.

The use of regularization improves the **clustering** of eigenvalues.

Influence of Regularization: `q25fv47`





HOPDM: Direct vs Iterative Methods

Problem	Dimensions		nonzeros(L)		
	nz(A)	nz(Q)	Direct	AS-Prec	NE-Prec
cvxqp1_m	1498	2984	75973	4739	4768
cvxqp1_l	14998	29984	3725045	71833	89241
cvxqp2_m	749	2984	51923	1031	315
cvxqp2_l	7499	29984	2754141	10579	3379
cvxqp3_m	2247	2984	90433	9527	14018
cvxqp3_l	22497	29984	4291057	149488	271780

QMR: Freund & Nachtigal (1991,1994).
QMR is asked for 10^{-3} accuracy.
500 MHz Pentium III PC, Linux, 256 MB.

Problem	Direct		AS-Prec		NE-Prec	
	its	time	its	time	its	time
cvxqp1_m	9	2.35	11	1.59	11	1.64
cvxqp1_l	11	1267.53	13	32.51	13	38.50
cvxqp2_m	9	1.27	10	1.01	10	1.06
cvxqp2_l	8	547.91	10	17.87	10	18.10
cvxqp3_m	9	3.40	11	1.94	11	2.37
cvxqp3_l	8	958.59	10	42.03	10	57.12

QMR: Freund & Nachtigal (1991,1994).
QMR is asked for 10^{-3} accuracy.
500 MHz Pentium III PC, Linux, 256 MB.

Problem	AS-Prec				NE-Prec			
	IPM	ItSl	Max	Avr	IPM	ItSl	Max	Avr
cvxqp1_m	11	338	20	14	11	338	20	14
cvxqp1_l	13	481	20	17	13	487	20	17
cvxqp2_m	10	307	20	13	10	307	20	13
cvxqp2_l	10	389	20	18	10	389	20	18
cvxqp3_m	11	303	20	13	11	296	20	12
cvxqp3_l	10	415	20	19	10	374	20	17

NL iterations (QMR) in the last IPM iteration:

Problem	AS-Prec		NE-Prec	
	Predictor	Corrector	Predictor	Corrector
cvxqp1_m	11	11	11	11
cvxqp1_l	16	12	14	12
cvxqp2_m	9	1	9	1
cvxqp2_l	18	16	18	16
cvxqp3_m	10	7	9	7
cvxqp3_l	20	14	15	13

GMRES: Saad & Schultz (1986).
BiCGSTAB: Van der Vorst (1992).
QMR: Freund & Nachtigal (1991,1994).

All approaches iterate until 10^{-3} accuracy is reached but perform no more than 20 iterations.
All approaches use the AS preconditioner.

500 MHz Pentium III PC, Linux, 256 MB.

Problem	GMRES			BiCGSTAB			QMR		
	IP	ItSl	time	IP	ItSl	time	IP	ItSl	time
cvxqp1_s	12	177	0.1	12	137	0.1	9	189	0.1
cvxqp1_m	11	307	1.3	11	233	1.3	11	338	1.6
cvxqp1_l	13	503	24.9	13	357	28.9	13	481	32.5
cvxqp2_s	27	217	0.1	16	153	0.1	10	235	0.1
cvxqp2_m	16	270	0.9	21	221	1.1	10	307	1.0
cvxqp2_l	19	404	16.1	10	243	14.9	10	389	18.1
cvxqp3_s	11	162	0.1	11	114	0.1	11	181	0.1
cvxqp3_m	11	306	1.6	11	226	1.7	11	303	1.9
cvxqp3_l	10	375	28.8	10	272	32.5	10	415	42.0

Conclusions

Direct Methods are reliable but occasionally excessively expensive.

Iterative Methods are promising but:

- are sometimes unpredictable;
- need tuning;
- depend upon preconditioners.

An **Augmented System** offers more freedom

- when used in the direct approach,
- when used to compute the preconditioners for the iterative approach.

Regularization is helpful.

What's to come in IPMs?

Direct Methods:

- small improvements:
  - reordering strategies
  - implementation (cache, supernodes)
- exploiting structure in huge problems (implicit inverse representations)

Iterative Methods:

- new preconditioners

Challenge:
Find an inverse representation with the number of nonzeros comparable to that of $\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$.