School of Mathematics

Practical Aspects of Large Scal Interior-Point Methods

Jacek Gondzio

Outline

Interior Point Methods:

- have been around for over 20 years...
- are competitive for small problems ($\leq$ 1,000,000 variables)
- are the only real approach for large problems ($\geq$ 1,000,000 variables)

Why are IPMs so efficient?

What can we do to improve them further?

Interior Point Methods

- **Fiacco & McCormick (1968)**
  handling inequality constraints - logarithmic barrier;
  minimization with inequality constraints
  replaced by a sequence of unconstrained minimizations
- **Lagrange (1788)**
  handling equality constraints - multipliers;
  minimization with equality constraints
  replaced by unconstrained minimization
- **Newton (1687)**
  solving unconstrained minimization problems;

Marsten, Subramanian, Saltzman, Lustig and Shanno:
"Interior point methods for linear programming:
Just call Newton, Lagrange, and Fiacco and McCormick!",
Interfaces 20 (1990) No 4, pp. 105–116.

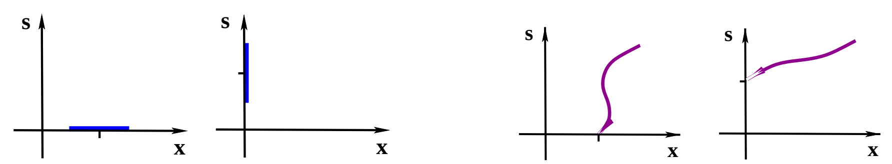
First Order Optimality Conditions

Simplex Method:

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= 0 \\ x,\, s &\geq 0. \end{aligned}$$

Interior Point Method:

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= \mu e \\ x,\, s &\geq 0. \end{aligned}$$



Basic: x > 0, s = 0 Nonbasic: x = 0, s > 0 "Basic": x > 0, s = 0 "Nonbasic": x = 0, s > 0

Theory: IPMs converge in $\mathcal{O}(\sqrt{n})$ or $\mathcal{O}(n)$ iterations
Practice: IPMs converge in $\mathcal{O}(\log n)$ iterations
... but one iteration may be expensive!

Direct Methods: Symmetric LDL^T Factorization

Indefinite	Quasidefinite	Positive Definite
$H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$	$H = \begin{bmatrix} Q & A^T \\ A & -R \end{bmatrix}$	$H = AQ^{-1}A^T$
2×2 pivots needed	1×1 pivots (any sign)	1×1 pivots (positive)
$\begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & a \\ a & d \end{bmatrix}$	strongly factorizable	easy

Vanderbei, *SIOPT* (1995): Symmetric QDFM's are strongly factorizable.
For any quasidefinite matrix there exists a **Cholesky-like** factorization

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

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

Minimum Degree Ordering

Sparse Matrix	Pivot h_{11}	Pivot h_{22}
$H = \begin{bmatrix} x & & x & x & x & \\ & x & & & x & \\ x & & x & & & x \\ x & & & x & & x \\ x & x & & & x & \\ & & x & x & & x \end{bmatrix}$	$\begin{bmatrix} \mathbf{p} & & \mathbf{x} & \mathbf{x} & \mathbf{x} & \\ & x & & & x & \\ \mathbf{x} & & x & \mathbf{f} & \mathbf{f} & x \\ \mathbf{x} & & \mathbf{f} & x & \mathbf{f} & x \\ \mathbf{x} & x & \mathbf{f} & \mathbf{f} & x & \\ & & x & x & & x \end{bmatrix}$	$\begin{bmatrix} x & & x & x & x & \\ & \mathbf{p} & & & \mathbf{x} & \\ x & & x & & & x \\ x & & & x & & x \\ x & \mathbf{x} & & & x & \\ & & x & x & & x \end{bmatrix}$

Minimum degree ordering:
choose a diagonal element corresponding to a row with the *min* number of nonzeros.
Permute rows and columns of H accordingly.

Optimality Conditions:

$$\begin{aligned} Ax &= b \\ A^Ty + s &= c \\ XSe &= \mu e \\ x,\ s &\geq 0. \end{aligned}$$

Newton Direction:

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

Linear Algebra involves an (ill-conditioned) scaling matrix $\Theta = XS^{-1}$.

Augmented System vs Normal Equations

LP	QP	NLP
$\begin{bmatrix} \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$	$\begin{bmatrix} Q+\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$	$\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$
$(A\Theta A^T)\Delta y = g$	$(A(Q+\Theta^{-1})^{-1}A^T)\Delta y = g$	$(AQ^{-1}A^T + ZY^{-1})\Delta y = g$

Theory of Interior Point Methods:

- very well understood for LP/QP problems
  Wright, "Primal-Dual Interior-Point Methods", SIAM, 1997.
- ongoing research on IPMs for NLP problems
  Nocedal & Wright, "Numerical Optimization", Springer, 1999.
  Conn, Gould & Toint, "Trust-Region Methods", SIAM, 2000.

Newton Liberation Front (Ph. Toint, 2004)

"Let the Newton method do the optimization"

in: Hager et al. (eds) *Multiscale Optimization Methods and Applications.*

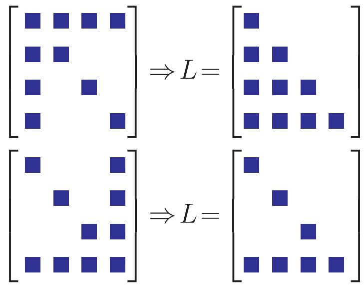
The rest of the talk
$\longrightarrow$ focuses on linear algebra issues.

From Sparsity to Block-Sparsity:

Sparse Matrix / **Block-Sparse Matrix**

$$H = \begin{bmatrix} x & x & x & x \\ x & x & & \\ x & & x & \\ x & & & x \end{bmatrix} \Rightarrow L = \begin{bmatrix} x & & & \\ x & x & & \\ x & x & x & \\ x & x & x & x \end{bmatrix}$$

$$PHP^T = \begin{bmatrix} x & & & x \\ & x & & x \\ & & x & x \\ x & x & x & x \end{bmatrix} \Rightarrow L = \begin{bmatrix} x & & & \\ & x & & \\ & & x & \\ x & x & x & x \end{bmatrix}$$



Object-Oriented Parallel Solver ⇒ problems of size $\mathbf{10^6}, \mathbf{10^7}, \mathbf{10^8}, \mathbf{10^9}, \ldots$
G. & Sarkissian, *MP* 96 (2003) 561-584.
G. & Grothey, *SIOPT* 13 (2003) 842-864.
G. & Grothey, *AOR* (to appear).

Talk of **Andreas Grothey** later today 4.00-4.25.

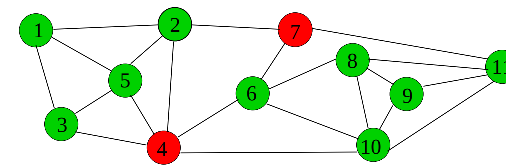
Inefficient Direct Approach

Cholesky factors get sometimes hopelessly dense.
QAP (Quadratic Assignment Problems).

Problem	Dimensions		
	rows	columns	nonzeros
qap12	3192	8856	38304
qap15	6330	22275	94950

Problem	**Normal Equations**			**Augmented System**		
	nz(AAt)	nz(LLt)	Flops	nz(A)	nz(LLt)	Flops
qap12	74592	2135388	2.378e+9	38304	1969957	2.046e+9
qap15	186075	8191638	1.792e+10	94950	7374972	1.522e+10

Nested Dissection:



Original Matrix

	1	2	3	4	5	6	7	8	9	10	11
1	x	x	x		x						
2	x	x		x	x		x				
3	x		x	x	x						
4		x	x	x	x	x				x	
5	x	x	x	x	x						
6				x		x	x	x		x	
7		x				x	x				x
8						x		x	x	x	x
9								x	x	x	x
10				x		x		x	x	x	x
11							x	x	x	x	x

Reordered Matrix

	1	2	3	5	6	8	9	10	11	4	7
1	x	x	x	x							
2	x	x		x						X	X
3	x		x	x						X	
5	x	x	x	x						X	
6					x	x		x		X	X
8					x	x	x	x	x		
9						x	x	x	x		
10					x	x	x	x	x	X	
11						x	x	x	x		X
4		X	X	X	X			X		X	
7		X			X				X		X

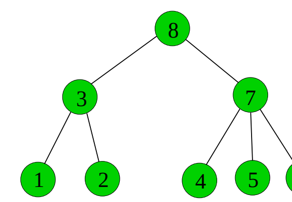
Matrix

$$\begin{bmatrix} x & & x & & & & & x \\ & x & x & & & & & x \\ x & x & x & & & & & x \\ & & & x & & & x & x \\ & & & & x & & x & x \\ & & & & & x & x & x \\ & & & x & x & x & x & x \\ x & x & x & x & x & x & x & x \end{bmatrix}$$

Cholesky Factor

$$\begin{bmatrix} x & & & & & & & \\ & x & & & & & & \\ x & x & x & & & & & \\ & & & x & & & & \\ & & & & x & & & \\ & & & & & x & & \\ & & & x & x & x & x & \\ x & x & x & x & x & x & x & x \end{bmatrix}$$

Elimination Tre



Supernodes

- small dense windows
- high level BLAS

$$\begin{bmatrix} x & & & & & & \\ x & x & & & & & \\ x & x & x & & & & \\ & & & \ddots & & & \\ x & x & x & & x & & \\ x & x & x & & & x & \\ x & x & x & & & & x \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} x & & & & & & \\ & x & & & & & \\ x & & x & & & & \\ & & & \ddots & & & \\ x & x & x & & x & & \\ x & x & x & & & x & \\ x & x & x & & & & x \end{bmatrix}$$

The Preconditioner $P = EE^T$ should:

- be easy to compute
  (significantly less expensive than Cholesky factor)
- be easy to invert
- produce good spectral properties of $E^{-1}HE^{-T}$ (that is $P^{-1}H$):
  either have few distinct eigenvalues;
  or have all eigenvalues in a small cluster: $\lambda_{min} \leq \lambda \leq \lambda_{max}$.

Examples:

- **Gill, Murray, Ponceleon & Saunders**, *SIMAX* 13 (1992) 292-311.
- **Murphy, Golub & Wathen**, *SISC* 21 (2000) 1969-1972.
- **Keller, Gould & Wathen**, *SIMAX* 21 (2000) 1300-1317.
  Gould, Hribal & Nocedal, *SISC* 23 (2001) 1376-1395.
- **Bergamaschi, G. & Zilli**, *COAP* 28 (2004) 149-171.
- **Golub & Grief**, *SISC* 24 (2003) 2076-2092;
  Grief, Golub & Varah, *SIMAX* (to appear).
- **Bai, Golub & Ng**, *SIMAX* 24 (2003) 603-626.

Gill, Murray, Ponceleón, Saunders, *SIMAX* 13 (1992) 292-311.
Compute Bunch-Parlett-Kaufmann factorization

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

where D is block-diagonal with 1×1 and 2×2 blocks.
Define the preconditioner $P = L\bar{D}L^T$, where $\bar{D}$ is a pdf approximation of D:

For 1×1 pivot:
replace d_{ii} by $\bar{d}_{ii} = |d_{ii}|$.

For 2×2 pivot:

replace $D_{i,i+1} = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} = \begin{bmatrix} c & s \\ s & -c \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} c & s \\ s & -c \end{bmatrix}$

by $\bar{D}_{i,i+1} = \begin{bmatrix} \bar{\alpha} & \bar{\beta} \\ \bar{\beta} & \bar{\gamma} \end{bmatrix} = \begin{bmatrix} c & s \\ s & -c \end{bmatrix} \begin{bmatrix} |\lambda_1| & \\ & |\lambda_2| \end{bmatrix} \begin{bmatrix} c & s \\ s & -c \end{bmatrix}$.

The preconditioned matrix has at most two distinct eigenvalues $+1$ and -1.
→ Use SYMMLQ (Paige and Saunders).

Iterative Methods

Normal Equations or **Augmented System**:

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

Oliveira *PhD*, Rice U., 1997; **Oliveira & Sorensen** *LAA* 394 (2005) 1-24.
→ It is better to precondition AS.
O, OS show that all preconditioners for the NE have an equivalent for the A
while the opposite is not true.
After all, NE is equivalent to a restricted order of pivoting in AS.

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}.$$

- Optimization: *KKT System*
- PDE: *Saddle Point Problem*

Benzi, Golub & Liesen, "Numerical Solution of Saddle Point Problems",
Acta Numerica 2005 (to appear).

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.

Rozlozník & Simoncini, *SIMAX* 24 (2002) 368-391.
RS consider the preconditioner P which guarantees that all eigenvalues of t
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
  that obtained for a positive definite matrix with the same eigenvalues as t
  original system.

How to choose G?

Bergamaschi, G. & Zilli, *COAP* 28 (2004) 149-171.
Augmented system in **QP, NLP**

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

Drop off-diagonal elements from Q:

Replace $\mathbf{Q} + \Theta^{-1}$ by $D = diag(\mathbf{Q}) + \Theta^{-1}$.

- With diagonal matrix D we have a choice between $\begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix}$ and $AD^{-1}A^T$.
- It is important to keep Θ^{-1} in the preconditioner. Θ is ill-conditioned:

For **"basic"** variables:	$\Theta_j = x_j/s_j \rightarrow \infty$	$\Theta_j^{-1} \rightarrow 0$;
For **"non-basic"** variables:	$\Theta_j = x_j/s_j \rightarrow 0$	$\Theta_j^{-1} \rightarrow \infty$.

Motivation: Sparsity issues: irreducible blocks in QP.

Consider the matrices

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

$$H = \left[\begin{array}{ccccc|cccc} \mathbf{x} & \mathbf{x} & & & & x & x & & \\ \mathbf{x} & \mathbf{x} & & & & & & x & x \\ & & x & & & & x & x & \\ & & & x & & & & & x \\ & & & & x & x & & x & \\ \hline x & & & & x & & & & \\ x & & x & & & & & & \\ & x & x & & x & & & & \\ & x & & x & & & & & \end{array}\right] \rightarrow H^{(2)} = \left[\begin{array}{ccccc|cccc} \mathbf{x} & \mathbf{x} & & & & x & x & \mathbf{f} & \mathbf{f} \\ \mathbf{x} & \mathbf{x} & & & & \mathbf{f} & \mathbf{f} & x & x \\ & & x & & & & x & x & \\ & & & x & & & & & x \\ & & & & x & x & & x & \\ \hline x & \mathbf{f} & & & x & \mathbf{f} & \mathbf{f} & \mathbf{f} & \mathbf{f} \\ x & \mathbf{f} & x & & & \mathbf{f} & \mathbf{f} & \mathbf{f} & \mathbf{f} \\ \mathbf{f} & x & x & & x & \mathbf{f} & \mathbf{f} & \mathbf{f} & \mathbf{f} \\ \mathbf{f} & x & & x & & \mathbf{f} & \mathbf{f} & \mathbf{f} & \mathbf{f} \end{array}\right].$$

Low Degree Minimum Polynomial

Murphy, Golub & Wathen, *SISC* 21 (2000) 1969-1972.
Consider 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 the preconditioner which incorporates an exact Schur complement AQ^{-}
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 ca
be very effective even though they are in no sense approximate inverses".

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:

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

where $G \in \mathcal{R}^{n \times n}$ 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 a
positive and satisfy

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

Primal and Dual Regularization

Primal Problem

$$\begin{array}{ll} \min & z_P = c^Tx + \frac{1}{2}x^TQx - \mu\sum_{j=1}^n \ln x_j \\ \text{s.t.} & Ax = b, \; x \geq 0 \end{array}$$

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

Primal Regularized Problem

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

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

Dual Problem

$$\begin{array}{ll} \max & z_D = b^Ty - \frac{1}{2}x^TQx + \mu\sum_{j=1}^n \ln s_j \\ \text{s.t.} & A^Ty + s - Qx = c, \\ & s \geq 0 \end{array}$$

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

Dual Regularized Problem

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

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

Augmented Lagrangian Regularization

Golub & Grief, *SISC* 24 (2003) 2076-2092;
Grief, Golub & Varah, *SIMAX* (to appear)
see also **Fletcher** (1975).

Replace $H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$ by $H_W = \begin{bmatrix} Q+A^TWA & A^T \\ A & 0 \end{bmatrix}$

Replace $\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$ by $\begin{bmatrix} Q+A^TWA & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f+A^TWd \\ d \end{bmatrix}$,

where W is a weight matrix, say, $W = \gamma I$.

Dostál & Schöberl, *COAP* 30 (2005) 23-43.
→ Use $Q + A^TWA$ only in matrix-vector multiplications.
Application to numerical solution of elliptic variational inequalities.

Spectral Analysis:

Eigenvalues of $P^{-1}H$ satisfy:

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

If $\lambda = 1$, we are done. If $\lambda \neq 1$ the second equation yields $Ax = 0$. After multiplying the first equation with x^T, we get:

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

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

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

and finally

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

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

Primal-Dual Regularization

Altman & G., *OMS* 11-12 (1999) 275-302.
Interpretation: proximal terms added to primal/dual objectives;
Dynamic regularization: correct only suspicious pivots.

Replace $H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$ by $H_R = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} + \begin{bmatrix} R_p & 0 \\ 0 & -R_d \end{bmatrix}$.

Replace $P = \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix}$ by $P_R = \begin{bmatrix} D & A^T \\ A & 0 \end{bmatrix} + \begin{bmatrix} R_p & 0 \\ 0 & -R_d \end{bmatrix}$.

Eigenvalues of the preconditioned matrix change:

$\lambda(P^{-1}H) = \dfrac{x^TQx}{x^TDx}$ is replaced by $\lambda(P_R^{-1}H_R) = \dfrac{x^TQx + \delta}{x^TDx + \delta}$,

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

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

Keller, Gould & Wathen, *SIMAX* 21 (2000) 1300-1317.
Gould, Hribar & Nocedal, *SISC* 23 (2001) 1376-1395.

Null space representation of A: given a basic/nonbasic partition $A = [B|N]$ with nonsingular B the columns of $Z = \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix}$ span null space of A.

Constraint Preconditioner

Replace $H = \left[\begin{array}{cc|c} Q_{BB}+\Theta_B^{-1} & Q_{BN} & B^T \\ Q_{NB} & Q_{NN}+\Theta_N^{-1} & N^T \\ \hline B & N & 0 \end{array}\right]$ by $P = \left[\begin{array}{cc|c} G_{BB} & G_{BN} & B^T \\ G_{NB} & G_{NN} & N^T \\ \hline B & N & 0 \end{array}\right]$

Many options:

- drop Q_{NB}, Q_{BN} (that is, set $G_{NB}=0$ and $G_{BN}=0$);
- replace $Q_{BB}+\Theta_B^{-1}$ by $G_{BB} = diag(Q_{BB}+\Theta_B^{-1})$;
- replace $Q_{NN}+\Theta_N^{-1}$ by $G_{NN} = diag(Q_{NN}+\Theta_N^{-1})$.

Dollar, Gould & Wathen, *RAL-TR-2004-036* (2004).

Two Options:

Option 1: $V = \begin{bmatrix} V_1 & V_2 \\ A & \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2^T \\ \Sigma_2 & \Sigma_3 \end{bmatrix}$

$$P = V\Sigma V^T = \begin{bmatrix} V_1\Sigma_1V_1^T + V_2\Sigma_2V_1^T + V_1\Sigma_2^TV_2^T + V_2\Sigma_3V_2^T & V_1\Sigma_1A^T + V_2\Sigma_2A^T \\ A\Sigma_1V_1^T + A\Sigma_2^TV_2^T & A\Sigma_1A^T \end{bmatrix}$$

Option 2: $U = \begin{bmatrix} U_1 & A^T \\ U_2 & \end{bmatrix}$, $\Lambda = \begin{bmatrix} \Lambda_1 & \Lambda_2^T \\ \Lambda_2 & \Lambda_3 \end{bmatrix}$

$$P = U\Lambda U^T = \begin{bmatrix} U_1\Lambda_1U_1^T + A^T\Lambda_2U_1^T + U_1\Lambda_2^TA + A^T\Lambda_3A & U_1\Lambda_1U_2^T + A^T\Lambda_2U_2^T \\ U_2\Lambda_1U_1^T + U_2\Lambda_2^TA & U_2\Lambda_1U_2^T \end{bmatrix}$$

Option 2 offers more flexibility in reproducing:

- (2,1) block equal to A; and
- (2,2) block equal to 0.

Skew-Hermitian Preconditioning

Bai, Golub & Ng, *SIMAX* 24 (2003) 603-626.

Replace $\begin{bmatrix} Q & A^T \\ A & -R_d \end{bmatrix}$ by $H = \begin{bmatrix} Q & A^T \\ -A & R_d \end{bmatrix}$.

Define: $\mathcal{H} = \frac{1}{2}(H+H^T) = \begin{bmatrix} Q & \\ & R_d \end{bmatrix}$ and $\mathcal{K} = \frac{1}{2}(H-H^T) = \begin{bmatrix} & A^T \\ -A & \end{bmatrix}$.

Two splittings:

$$\begin{aligned} H = \mathcal{H}+\mathcal{K} &= (\mathcal{H}+\alpha I) - (\alpha I - \mathcal{K}), \\ H = \mathcal{H}+\mathcal{K} &= (\mathcal{K}+\alpha I) - (\alpha I - \mathcal{H}). \end{aligned}$$

Stationary iteration alternating between these two splittings:

$$\begin{aligned} (\mathcal{H}+\alpha I)v &= (\alpha I - \mathcal{K})u_k + b \\ (\mathcal{K}+\alpha I)u_{k+1} &= (\alpha I - \mathcal{H})v \; + b. \end{aligned}$$

After eliminating the intermediate variable v we get

$$u_{k+1} = \mathcal{T}_\alpha u_k + g,$$

where

$$\mathcal{T}_\alpha = (\mathcal{K}+\alpha I)^{-1}(\alpha I - \mathcal{H})(\mathcal{H}+\alpha I)^{-1}(\alpha I - \mathcal{K}).$$

An alternative *correction form*:

$$u_{k+1} = u_k + P_\alpha^{-1} r_k \qquad (r_k = b - Hu_k),$$

with the preconditioner

$$P_\alpha = \frac{1}{2\alpha}(\mathcal{H}+\alpha I)(\mathcal{K}+\alpha I).$$

Inversions of the regularized matrices are needed:

$$\mathcal{H}+\alpha I = \begin{bmatrix} Q & \\ & R_d \end{bmatrix} + \alpha I \quad \text{and} \quad \mathcal{K}+\alpha I = \begin{bmatrix} & A^T \\ -A & \end{bmatrix} + \alpha I.$$

Worry: it may be difficult to satisfy constraints with this preconditioner.
→ Thorough computational study needed.

Conclusions:

Direct Methods are reliable and well-suited to structure exploitation
but occasionally get excessively expensive.

Iterative Methods are promising
but need tuning and depend upon preconditioners.

What do we need?

- new inverse representation
- new preconditioners

Ultimate Objective

Find an inverse of $\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$ **with** $\mathcal{O}(nzQ) + \mathcal{O}(nzA)$ **nonzeros.**