Part 4: Active-set methods for linearly constrained optimization

Nick Gould (RAL)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \;\; \text{subject to} \;\; Ax \geq b$$

Part C course on continuoue optimization

LINEARLY CONSTRAINED MINIMIZATION

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; f(x) \;\; \text{subject to} \;\; Ax \left\{ \begin{array}{c} \geq \\ = \end{array} \right\} b$$

where the **objective function** $f : \mathbb{R}^n \longrightarrow \mathbb{R}$

- assume that $f \in C^1$ (sometimes C^2) and Lipschitz
- often in practice this assumption violated, but not necessary
- important special cases:
  - **linear programming**: $f(x) = g^T x$
  - **quadratic programming**: $f(x) = g^T x + \frac{1}{2} x^T H x$

Concentrate here on quadratic programming

QUADRATIC PROGRAMMING

QP: $$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; q(x) = g^T x + \tfrac{1}{2} x^T H x \;\; \text{subject to} \;\; Ax \geq b$$

- H is n by n, real symmetric, $g \in \mathbb{R}^n$
- $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$ is m by n real, $b = \begin{pmatrix} [b]_1 \\ \vdots \\ [b]_m \end{pmatrix}$
- in general, constraints may
  - have upper bounds: $b^l \leq Ax \leq b^u$
  - include equalities: $A^e x = b^e$
  - involve simple bounds: $x^l \leq x \leq x^u$
  - include network constraints ...

PROBLEM TYPES

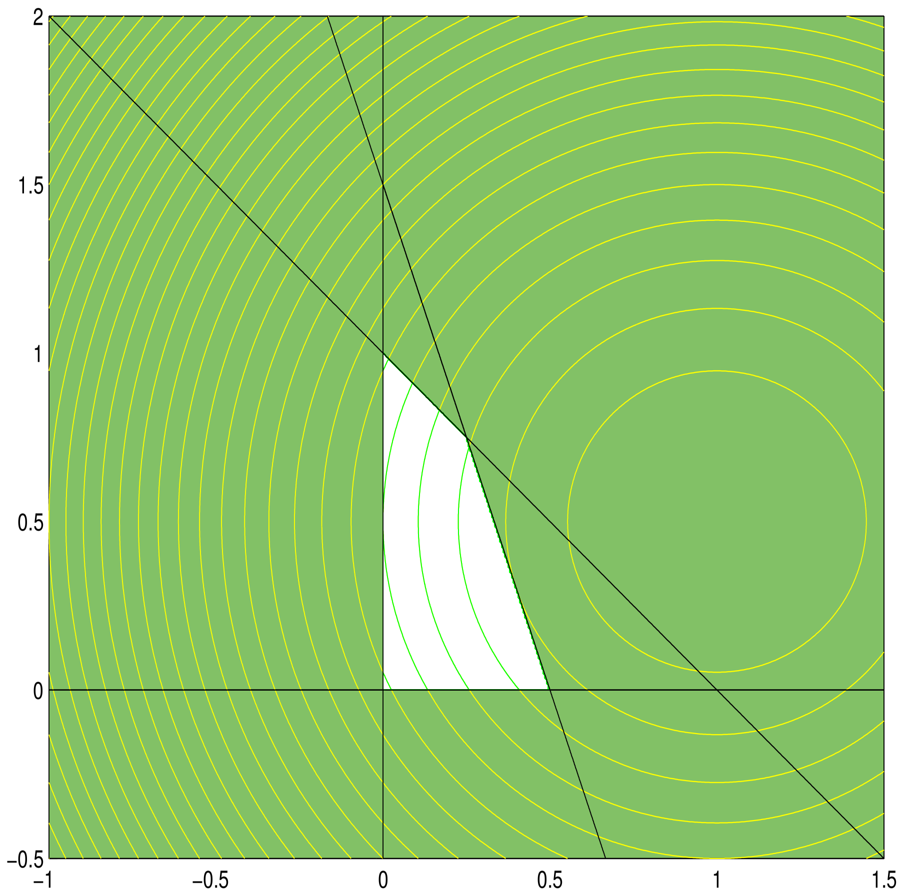
Convex problems

- H is positive semi-definite ($x^T H x \geq 0$ for all x)
- any local minimizer is global
- important special case: $H = 0 \iff$ linear programming

Strictly convex problems

- H is positive definite ($x^T H x > 0$ for all $x \neq 0$)
- unique minimizer (if any)

CONVEX EXAMPLE



Contours of objective function

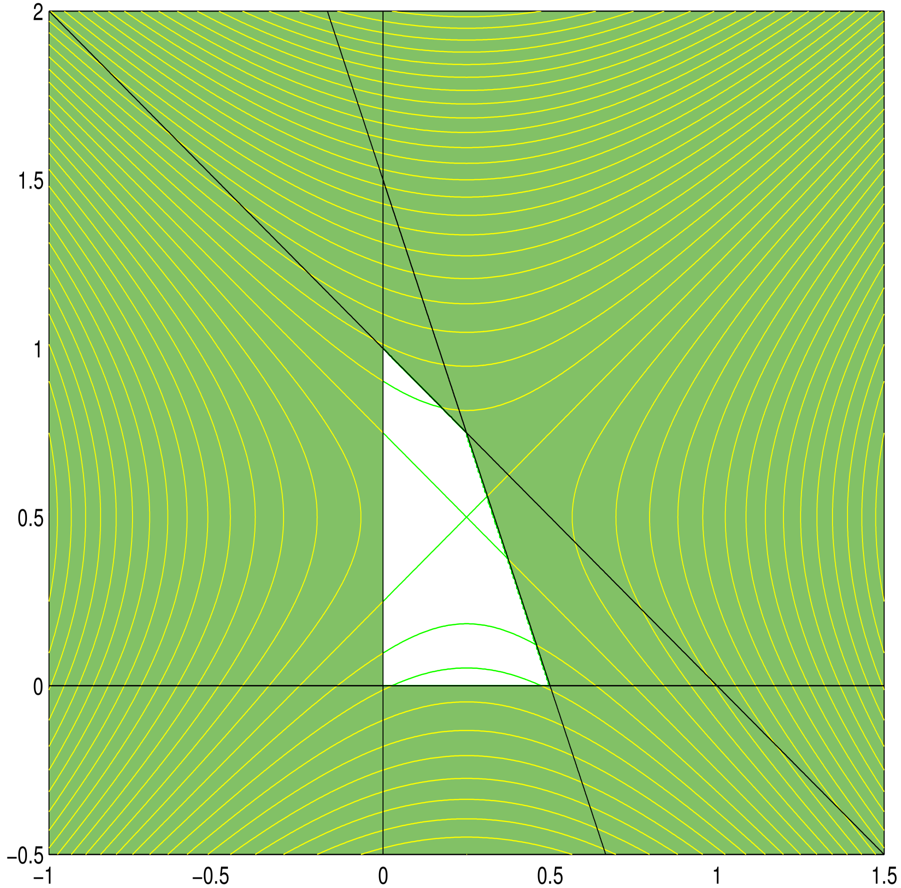
$$\min (x_1 - 1)^2 + (x_2 - 0.5)^2$$
$$\text{subject to} \;\; x_1 + x_2 \le 1$$
$$3x_1 + x_2 \le 1.5$$
$$(x_1, x_2) \ge 0$$

PROBLEM TYPES (II)

General (non-convex) problems

- H may be indefinite ($x^T H x < 0$ for some x)
- may be many local minimizers
- may have to be content with a local minimizer
- problem may be unbounded from below

NON-CONVEX EXAMPLE



Contours of objective function

$$
\begin{aligned}
\min\ & -2(x_1 - 0.25)^2 + 2(x_2 - 0.5)^2 \\
\text{subject to}\ & x_1 + x_2 \le 1 \\
& 3x_1 + x_2 \le 1.5 \\
& (x_1, x_2) \ge 0
\end{aligned}
$$

PROBLEM TYPES (III)

Small

- values/structure of matrix data H and A irrelevant
- currently $\min(m, n) = O(10^2)$

Large

- values/structure of matrix data H and A important
- currently $\min(m, n) \geq O(10^3)$

Huge

- factorizations involving H and A are unrealistic
- currently $\min(m, n) \geq O(10^5)$

WHY IS QP SO IMPORTANT?

- many **applications**
  - portfolio analysis, structural analysis, VLSI design, discrete-time stabilization, optimal and fuzzy control, finite impulse response design, optimal power flow, economic dispatch …
  - $\sim$ 500 application papers
- **prototypical** nonlinear programming problem
- **basic subproblem** in constrained optimization:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) \geq 0 \end{array} \quad \Longrightarrow \quad \begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f + g^T x + \frac{1}{2} x^T H x \\ \text{subject to} & Ax + c \geq 0 \end{array}$$

$\Longrightarrow$ SQP methods ($\Longrightarrow$ Course Part 7)

OPTIMALITY CONDITIONS

Recall: the importance of optimality conditions is:

- to be able to recognise a solution if found by accident or design
- to guide the development of algorithms

FIRST-ORDER OPTIMALITY

$$\text{QP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \; q(x) = g^T x + \tfrac{1}{2} x^T H x \; \text{ subject to } \; Ax \geq b$$

Any point x_* that satisfies the conditions

$$Ax_* \geq b \qquad (\textbf{primal feasibility})$$

$$Hx_* + g - A^T y_* = 0 \;\text{ and }\; y_* \geq 0 \qquad (\textbf{dual feasibility})$$

$$[Ax_* - b]_i \cdot [y_*]_i = 0 \;\text{ for all }\; i \qquad (\textbf{complementary slackness})$$

for some vector of **Lagrange multipliers** y_* is a **first-order critical** (or Karush-Kuhn-Tucker) point

If $[Ax_* - b]_i = 0 \iff [y_*]_i > 0$ for all $i \Longrightarrow$ the solution is **strictly complementary**

SECOND-ORDER OPTIMALITY

$$\text{QP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \; q(x) = g^T x + \tfrac{1}{2} x^T H x \; \text{ subject to } \; Ax \geq b$$

Let

$$\mathcal{N}_+ = \left\{ s \;\middle|\; \begin{array}{l} a_i^T s = 0 \text{ for all } i \text{ such that } a_i^T x_* = [b]_i \text{ and } [y_*]_i > 0 \text{ and} \\ a_i^T s \geq 0 \text{ for all } i \text{ such that } a_i^T x_* = [b]_i \text{ and } [y_*]_i = 0 \end{array} \right\}$$

Any first-order critical point x_* for which additionally

$$s^T H s \geq 0 \;\; (\text{resp.} > 0) \;\; \text{for all} \;\; s \in \mathcal{N}_+$$

is a **second-order** (resp. **strong second-order**) critical point

Theorem 4.1: x_* is a (an isolated) local minimizer of QP $\Longleftrightarrow$ x_* is (strong) second-order critical

WEAK SECOND-ORDER OPTIMALITY

$$\text{QP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \; q(x) = g^T x + \tfrac{1}{2} x^T H x \; \text{ subject to } \; Ax \geq b$$

Let

$$\mathcal{N} = \left\{ s \mid a_i^T s = 0 \;\text{ for all }\; i \;\text{ such that }\; a_i^T x_* = [b]_i \right\}$$

Any first-order critical point x_* for which additionally

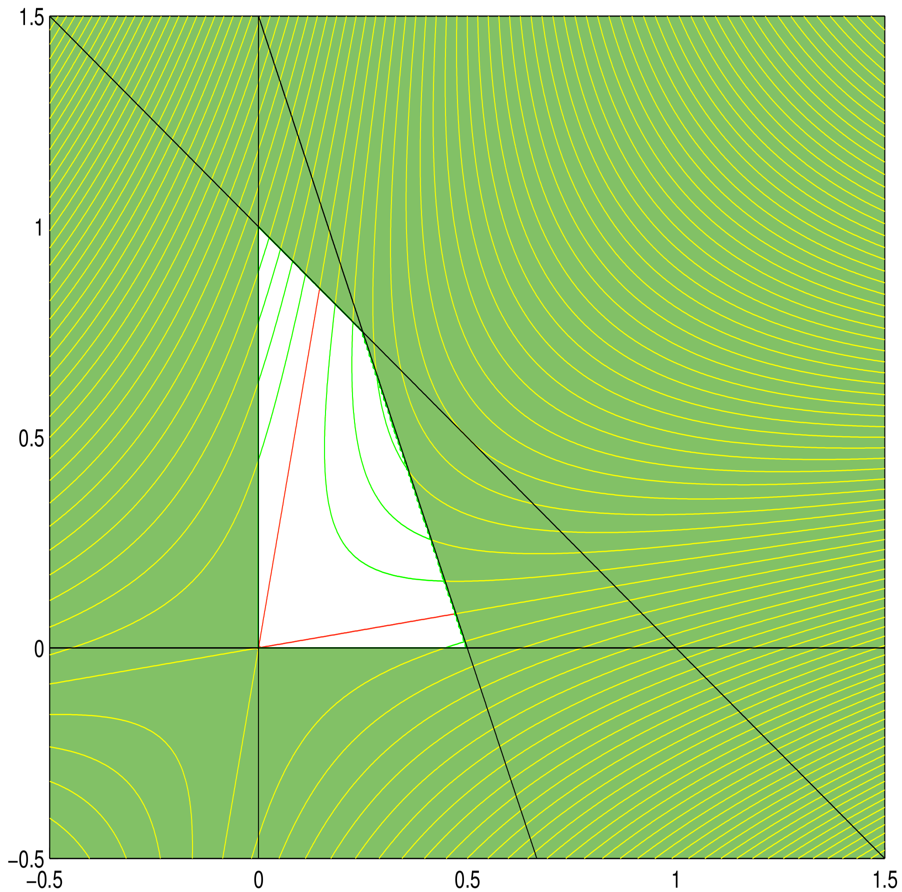
$$s^T H s \geq 0 \;\text{ for all }\; s \in \mathcal{N}$$

is a **weak** second-order critical point

Note that

- ⊙ a weak second-order critical point may be a maximizer!
- ⊙ checking for weak second-order criticality is easy (strong is hard)

NON-CONVEX EXAMPLE



$$\min x_1^2 + x_2^2 - 6x_1x_2$$
$$\text{subject to } \; x_1 + x_2 \le 1$$
$$3x_1 + x_2 \le 1.5$$
$$(x_1, x_2) \ge 0$$

Contours of objective function:

note that escaping from the origin may be difficult!

[DUALITY

$$\text{QP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; q(x) = g^T x + \tfrac{1}{2} x^T H x \;\; \text{subject to} \;\; Ax \geq b$$

If QP is convex, any first-order critical point is a global minimizer

If H is strictly convex, the problem

$$\underset{y \in \mathbb{R}^m,\, y \geq 0}{\text{maximize}} \quad -\tfrac{1}{2} g^T H^{-1} g + (AH^{-1} g + b)^T y - \tfrac{1}{2} y^T A H^{-1} A^T y$$

is known as the **dual** of QP

- ⊙ QP is the **primal**
- ⊙ primal and dual have same KKT conditions
- ⊙ if primal is feasible, optimal value of primal = optimal value dual
- ⊙ can be generalized for simply convex case]

ALGORITHMS

Essentially two classes of methods (slight simplification)

active set methods :

- **primal** active set methods aim for dual feasibility while maintaining primal feasibility and complementary slackness
- **dual** active set methods aim for primal feasibility while maintaining dual feasibility and complementary slackness

interior-point methods : aim for complementary slackness while maintaining primal and dual feasibility ($\Longrightarrow$ Course Part 6)

EQUALITY CONSTRAINED QP

The basic subproblem in all of the methods we will consider is

EQP: $\underset{x \in \mathbb{R}^n}{\text{minimize}} \; g^T x + \frac{1}{2} x^T H x \; \text{ subject to } \; Ax = 0 \longleftarrow$ **N.B.**

Assume A is m by n, full-rank (preprocess if necessary)

- ⊙ First-order optimality (Lagrange multipliers y)

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}$$

- ⊙ Second-order necessary optimality:
  $s^T H s \geq 0$ for all s for which $As = 0$

- ⊙ Second-order sufficient optimality:
  $s^T H s > 0$ for all $s \neq 0$ for which $As = 0$

EQUALITY CONSTRAINED QP (II)

$$\text{EQP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; q(x) = g^T x + \tfrac{1}{2} x^T H x \;\; \text{subject to} \;\; Ax = 0$$

Four possibilities:

(i)
$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix} \qquad (*)$$

and H is second-order sufficient $\Longrightarrow$ **unique** minimizer x

(ii) $(*)$ holds, H is second-order necessary, but $\exists s$ such that $Hs = 0$ and $As = 0 \Longrightarrow$ family of **weak** minimizers $x + \alpha s$ for any $\alpha \in \mathbb{R}$

(iii) $\exists s$ for which $As = 0$, $Hs = 0$ and $g^T s < 0 \Longrightarrow$ $q(\cdot)$ unbounded along **direction of linear infinite descent** s

(iv) $\exists s$ for which $As = 0$ and $s^T H s < 0 \Longrightarrow$ $q(\cdot)$ unbounded along **direction of negative curvature** s

CLASSIFICATION OF EQP METHODS

Aim to solve

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}$$

Three basic approaches:

- **full-space** approach
- **range-space** approach
- **null-space** approach

For each of these can use

- **direct** (factorization) method
- **iterative** (conjugate-gradient) method

FULL-SPACE/KKT/AUGMENTED SYSTEM APPROACH

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}$$

- **KKT matrix**

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$$

  is symmetric, indefinite $\Longrightarrow$ use Bunch-Parlett type factorization
  - $K = PLBL^TP^T$
  - P permutation, L unit lower-triangular
  - B block diagonal with 1x1 and 2x2 blocks
- LAPACK for small problems, `MA27`/`MA57` for large ones
- **Theorem 4.2**: H is second-order sufficient $\Longleftrightarrow$
  K non-singular and has precisely m negative eigenvalues

RANGE-SPACE APPROACH

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix} \qquad (*)$$

For **non-singular** H

- ⊙ eliminate x using first block of $(*)$ $\Longrightarrow$

$$AH^{-1}A^T y = AH^{-1} g \;\text{ followed by }\; Hx = -g + A^T y$$

- ⊙ strictly convex case $\Longrightarrow$ H and $AH^{-1}A^T$ positive definite $\Longrightarrow$ Cholesky factorization
- ⊙ **Theorem 4.3**: H is second-order sufficient $\Longleftrightarrow$ H and $AH^{-1}A^T$ have same number of negative eigenvalues
- ⊙ $AH^{-1}A^T$ usually dense $\Longrightarrow$ factorization only for small m

NULL-SPACE APPROACH

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix} \qquad (*)$$

- let n by $n-m$ S be a **basis** for null-space of $A \Longrightarrow AS = 0$
- second block $(*) \Longrightarrow x = Sx_N$
- premultiply first block $(*)$ by $S^T \Longrightarrow$

$$S^T H S x_S = -S^T g$$

- **Theorem 4.4**: H is second-order sufficient $\Longleftrightarrow$ $S^T H S$ is positive definite $\Longrightarrow$ Cholesky factorization
- $S^T H S$ usually dense $\Longrightarrow$ factorization only for small $n-m$

NULL-SPACE BASIS

Require n by $n-m$ null-space basis S for $A \Longrightarrow AS = 0$

Non-orthogonal basis: let $A = (A_1 \;\; A_2)P$

- P permutation, A_1 non-singular

$$\Longrightarrow S = P^T \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix}$$

- generally suitable for large problems. Best A_1?

Orthogonal basis: let $A = (L \;\; 0)Q$

- L non-singular (e.g., triangular), $Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$ orthonormal

$$\Longrightarrow S = Q_2^T$$

- more stable but . . . generally unsuitable for large problems

[ITERATIVE METHODS FOR SYMMETRIC LINEAR SYSTEMS

$$Bx = b$$

Best methods are based on finding solutions from the **Krylov space**

$$\mathcal{K} = \{r^0, Br^0, B(Br^0), \ldots\} \qquad (r^0 = b - Bx^0)$$

B **indefinite:** use MINRES method

B **positive definite:** use conjugate gradient method

- ⊙ usually satisfactory to find approximation rather than exact solution
- ⊙ usually try to **precondition** system, i.e., solve

$$C^{-1}Bx = C^{-1}b$$

where $C^{-1}B \approx I$]

[ITERATIVE RANGE-SPACE APPROACH

$$AH^{-1}A^Ty = AH^{-1}g \ \text{ followed by } \ Hx = -g + A^Ty$$

For strictly convex case $\Longrightarrow$ H and $AH^{-1}A^T$ positive definite

H^{-1} **available:** (directly or via factors),

use conjugate gradients to solve $AH^{-1}A^Ty = AH^{-1}g$

- ⊙ matrix vector product $AH^{-1}A^Tv = \left(A\left(H^{-1}(A^Tv)\right)\right)$
- ⊙ preconditioning? Need to approximate (likely dense) $AH^{-1}A^T$

H^{-1} **not available:** use composite conjugate gradient method (Urzawa's method) iterating both on solutions to

$$AH^{-1}A^Ty = AH^{-1}g \quad \text{and} \quad Hx = -g + A^Ty$$

at the same time (may not converge)]

ITERATIVE NULL-SPACE APPROACH

$$S^T H S x_N = -S^T g \text{ followed by } x = S x_N$$

- use conjugate gradient method
  - matrix vector product $S^T H S v_N = \left(S^T \left(H(S v_N)\right)\right)$
  - preconditioning? Need to approximate (likely dense) $S^T H S$
  - if we encounter s_N such that $s_N^T (S^T H S) s_N < 0 \Longrightarrow s = N s_N$ is a direction of negative curvature since $As = 0$ and $s^T H s < 0$
  - **Advantage**: $Ax^{\text{approx}} = 0$

ITERATIVE FULL-SPACE APPROACH

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix}$$

⊙ use MINRES with the preconditioner

$$\begin{pmatrix} M & 0 \\ 0 & AN^{-1}A^T \end{pmatrix}$$

where M and $N \approx H$.

- ⋄ **Disadvantage**: $Ax^{\text{approx}} \neq 0$

⊙ use conjugate gradients with the preconditioner

$$\begin{pmatrix} M & A^T \\ A & 0 \end{pmatrix}$$

where $M \approx H$.

- ⋄ **Advantage**: $Ax^{\text{approx}} = 0$

ACTIVE SET ALGORITHMS

$$\text{QP:} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; q(x) = g^T x + \tfrac{1}{2} x^T H x \;\; \text{subject to} \;\; Ax \geq b$$

The **active set** $\mathcal{A}(x)$ at x is

$$\mathcal{A}(x) = \{ i \mid a_i^T x = [b]_i \}$$

If x_* solves QP, we have

$$\begin{aligned} & \arg\min q(x) \;\; \text{subject to} \;\; Ax \geq b \\ \equiv \; & \arg\min q(x) \;\; \text{subject to} \;\; a_i^T x = [b]_i \;\; \text{for all} \;\; i \in \mathcal{A}(x_*) \end{aligned}$$

A **working set** $\mathcal{W}(x)$ at x is a **subset of the active set** for which the vectors $\{a_i\}$, $i \in \mathcal{W}(x)$ are linearly independent

BASICS OF ACTIVE SET ALGORITHMS

Basic idea: Pick a subset $\mathcal{W}_k$ of $\{1, \ldots, m\}$ and find

$$x_{k+1} = \arg\min q(x) \;\text{ subject to }\; a_i^T x = [b]_i \;\text{ for all }\; i \in \mathcal{W}_k$$

If x_{k+1} does not solve QP, adjust $\mathcal{W}_k$ to form $\mathcal{W}_{k+1}$ and repeat

Important issues are:

- ⊙ how do we know if x_{k+1} solves QP ?
- ⊙ if x_{k+1} does not solve QP, how do we pick the next working set $\mathcal{W}_{k+1}$?

Notation: rows of A_k are those of A indexed by $\mathcal{W}_k$
components of b_k are those of b indexed by $\mathcal{W}_k$

PRIMAL ACTIVE SET ALGORITHMS

Important feature: ensure all iterates are feasible, i.e., $Ax_k \geq b$

If $\mathcal{W}_k \subseteq \mathcal{A}(x_k)$
$\Longrightarrow A_k x_k = b_k$ and $A_k x_{k+1} = b_k$
$\Longrightarrow x_{k+1} = x_k + s_k$, where

$$\begin{aligned} s_k &= \arg\min \mathrm{EQP}_k \\ &= \arg \underbrace{\min q(x_k + s) \ \text{ subject to } \ A_k s = 0}_{\text{equality constrained problem}} \end{aligned}$$

Need an initial feasible point x_0

PRIMAL ACTIVE SET ALGORITHMS — ADDING CONSTRAINTS

$$s_k = \arg\min q(x_k + s) \;\text{ subject to }\; A_k s = 0$$

What if $x_k + s_k$ is not feasible?

- ⊙ a currently inactive constraint j must become active at $x_k + \alpha_k s_k$ for some $\alpha_k < 1$ — pick the smallest such α_k
- ⊙ move instead to $x_{k+1} = x_k + \alpha_k s_k$ and set $\mathcal{W}_{k+1} = \mathcal{W}_k + \{j\}$

PRIMAL ACTIVE SET ALGORITHMS — DELETING CONSTRAINTS

What if $x_{k+1} = x_k + s_k$ is feasible ? $\Longrightarrow$

$$x_{k+1} = \arg\min q(x) \;\text{ subject to }\; a_i^T x = [b]_i \;\text{ for all }\; i \in \mathcal{W}_k$$

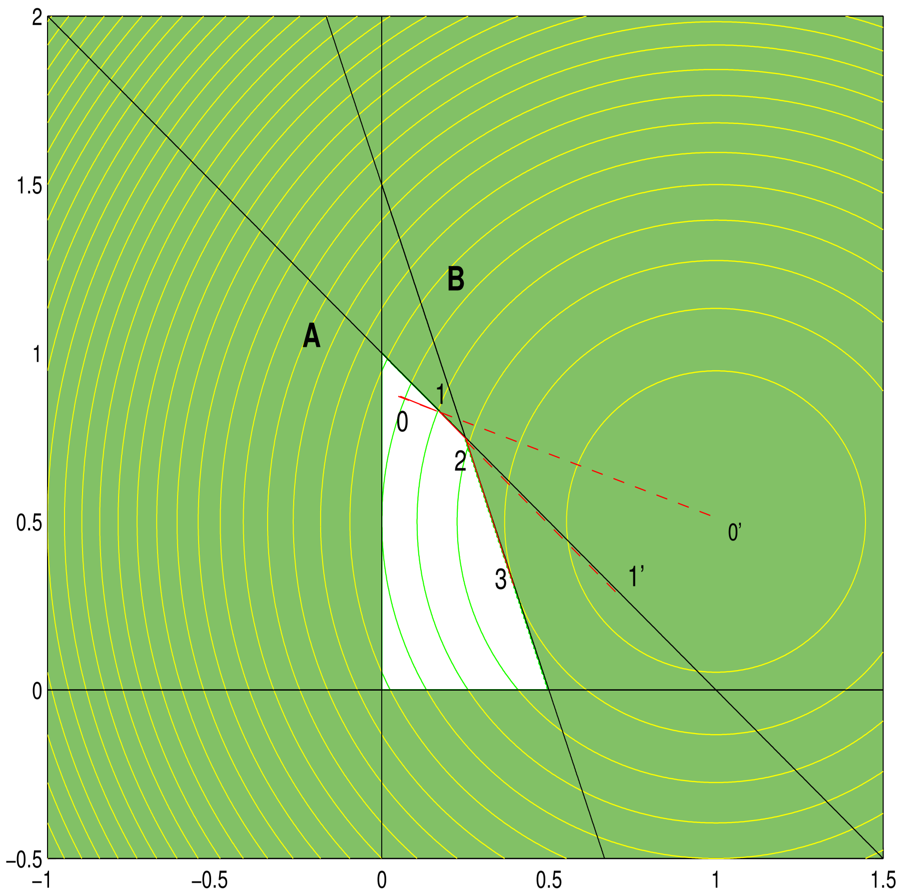
$\Longrightarrow$ $\exists$ Lagrange multipliers y_{k+1} such that

$$\begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} x_{k+1} \\ -y_{k+1} \end{pmatrix} = \begin{pmatrix} -g \\ b_k \end{pmatrix}$$

Three possibilities:

- $\odot$ $q(x_{k+1}) = -\infty$ (not strictly-convex case only)
- $\odot$ $y_{k+1} \geq 0 \Longrightarrow x_{k+1}$ is a first-order critical point of QP
- $\odot$ $[y_{k+1}]_i < 0$ for some $i \Longrightarrow q(x)$ may be improved by considering $\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{j\}$, where j is the i-th member of $\mathcal{W}_k$

ACTIVE-SET APPROACH



0. Starting point
0'. Unconstrained minimizer
1. Encounter constraint A
1'. Minimizer on constraint A
2. Encounter constraint B, move off constraint A
3. Minimizer on constraint B = required solution

LINEAR ALGEBRA

Need to solve a sequence of EQP_ks in which

$$\text{either} \quad \mathcal{W}_{k+1} = \mathcal{W}_k + \{j\} \quad \Longrightarrow \quad A_{k+1} = \begin{pmatrix} A_k \\ a_j^T \end{pmatrix}$$

$$\text{or} \quad \mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{j\} \quad \Longrightarrow \quad A_k = \begin{pmatrix} A_{k+1} \\ a_j^T \end{pmatrix}$$

Since working sets change gradually, aim to **update** factorizations rather than compute afresh

RANGE-SPACE APPROACH — MATRIX UPDATES

Need factors $L_{k+1}L_{k+1}^T = A_{k+1}H^{-1}A_{k+1}^T$ given $L_kL_k^T = A_kH^{-1}A_k^T$

When $A_{k+1} = \begin{pmatrix} A_k \\ a_j^T \end{pmatrix} \Longrightarrow$

$$A_{k+1}H^{-1}A_{k+1}^T = \begin{pmatrix} A_kH^{-1}A_k^T & A_kH^{-1}a_j \\ a_j^TH^{-1}A_k^T & a_j^TH^{-1}a_j \end{pmatrix}$$

$\Longrightarrow$

$$L_{k+1} = \begin{pmatrix} L_k & 0 \\ l^T & \lambda \end{pmatrix}$$

where

$$L_kl = A_kH^{-1}a_j \ \text{ and } \ \lambda = \sqrt{a_j^TH^{-1}a_j - l^Tl}$$

Essentially reverse this to remove a constraint

NULL-SPACE APPROACH — MATRIX UPDATES

Need factors $A_{k+1} = (L_{k+1} \;\; 0) Q_{k+1}$ given

$$A_k = (L_k \quad 0) Q_k = (L_k \;\; 0) \begin{pmatrix} Q_{1\,k} \\ Q_{2\,k} \end{pmatrix}$$

To add a constraint (to remove is similar)

$$\begin{aligned} A_{k+1} &= \begin{pmatrix} A_k \\ a_j^T \end{pmatrix} = \begin{pmatrix} L_k & 0 \\ a_j^T Q_{1\,k}^T & a_j^T Q_{2\,k}^T \end{pmatrix} Q_k \\ &= \begin{pmatrix} L_k & 0 \\ a_j^T Q_{1\,k}^T & a_j^T Q_{2\,k}^T \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & U^T \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & U \end{pmatrix} Q_k \\ &= \underbrace{\left[\begin{pmatrix} L_k & 0 \\ a_j^T Q_{1\,k}^T & \sigma e_1^T \end{pmatrix} \right]}_{(L_{k+1} \;\; 0)} \underbrace{\left[\begin{pmatrix} I & 0 \\ 0 & U \end{pmatrix} Q_k \right]}_{Q_{k+1}} \end{aligned}$$

where the Householder matrix U reduces $Q_{2\,k} a_j$ to $\sigma e_1 = \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$

[FULL-SPACE APPROACH — MATRIX UPDATES

$\mathcal{W}_k$ becomes $\mathcal{W}_\ell \Longrightarrow A_k = \begin{pmatrix} A_C \\ A_D \end{pmatrix}$ becomes $A_\ell = \begin{pmatrix} A_C \\ A_A \end{pmatrix}$

Solving

$$\begin{pmatrix} H & A_\ell^T \\ A_\ell & 0 \end{pmatrix} \begin{pmatrix} s_\ell \\ -y_\ell \end{pmatrix} = \begin{pmatrix} g_\ell \\ 0 \end{pmatrix} \Longrightarrow$$

$$\begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix} \longleftarrow \begin{pmatrix} \boxed{\begin{matrix} H & A_C^T & A_D^T \\ A_C & 0 & 0 \\ A_D & 0 & 0 \end{matrix}} & \begin{matrix} A_A^T & 0 \\ 0 & 0 \\ 0 & I \end{matrix} \\ \begin{matrix} A_A & 0 & 0 \\ 0 & 0 & I \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \end{pmatrix} \begin{pmatrix} s_\ell \\ -y_C \\ -y_D \\ -y_A \\ u_\ell \end{pmatrix} = \begin{pmatrix} g_\ell \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};$$

$$y_\ell = \begin{pmatrix} y_C \\ y_A \end{pmatrix}$$

. . .]

[FULL-SPACE APPROACH — MATRIX UPDATES (CONT.)

…can solve

$$\begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix} \longleftarrow \begin{pmatrix} \boxed{\begin{matrix} H & A_C^T & A_D^T \\ A_C & 0 & 0 \\ A_D & 0 & 0 \end{matrix}} & \begin{matrix} A_A^T & 0 \\ 0 & 0 \\ 0 & I \end{matrix} \\ \begin{matrix} A_A & 0 & 0 \\ 0 & 0 & I \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \end{pmatrix} \begin{pmatrix} s_\ell \\ -y_C \\ -y_D \\ -y_A \\ u_\ell \end{pmatrix} = \begin{pmatrix} g_\ell \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

using the factors of

$$K_k = \begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix}$$

and the **Schur complement**

$$S_\ell = -\begin{pmatrix} A_A & 0 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix}^{-1} \begin{pmatrix} A_A^T & 0 \\ 0 & 0 \\ 0 & I \end{pmatrix}$$

…]

SCHUR COMPLEMENT UPDATING

- Major iteration starts with factorization of

$$K_k = \begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix}$$

- As $\mathcal{W}_k$ changes to $\mathcal{W}_\ell$, factorization of

$$S_\ell = -\begin{pmatrix} A_A & 0 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} H & A_k^T \\ A_k & 0 \end{pmatrix}^{-1} \begin{pmatrix} A_A^T & 0 \\ 0 & 0 \\ 0 & I \end{pmatrix}$$

  is **updated** not recomputed

- Once dim S_ℓ exceeds a given threshold, or it is cheaper to factorize/use K_ℓ than maintain/use K_k and S_ℓ, start the next major iteration

PHASE-1

To find an initial feasible point x_0 such that $Ax_0 \geq b$

- ⊙ use traditional (simplex) phase-1, or
- ⊙ let $r = \min(b - Ax_{\text{guess}}, 0)$, and solve $\quad [(x_0, \xi_0) = (x_{\text{guess}}, 1)]$

$$\underset{x \in \mathbb{R}^n,\, \xi \in \mathbb{R}}{\text{minimize}} \;\; \xi \;\; \text{subject to} \;\; Ax + \xi r \geq b \;\; \text{and} \;\; \xi \geq 0$$

Alternatively, use a single-phase method

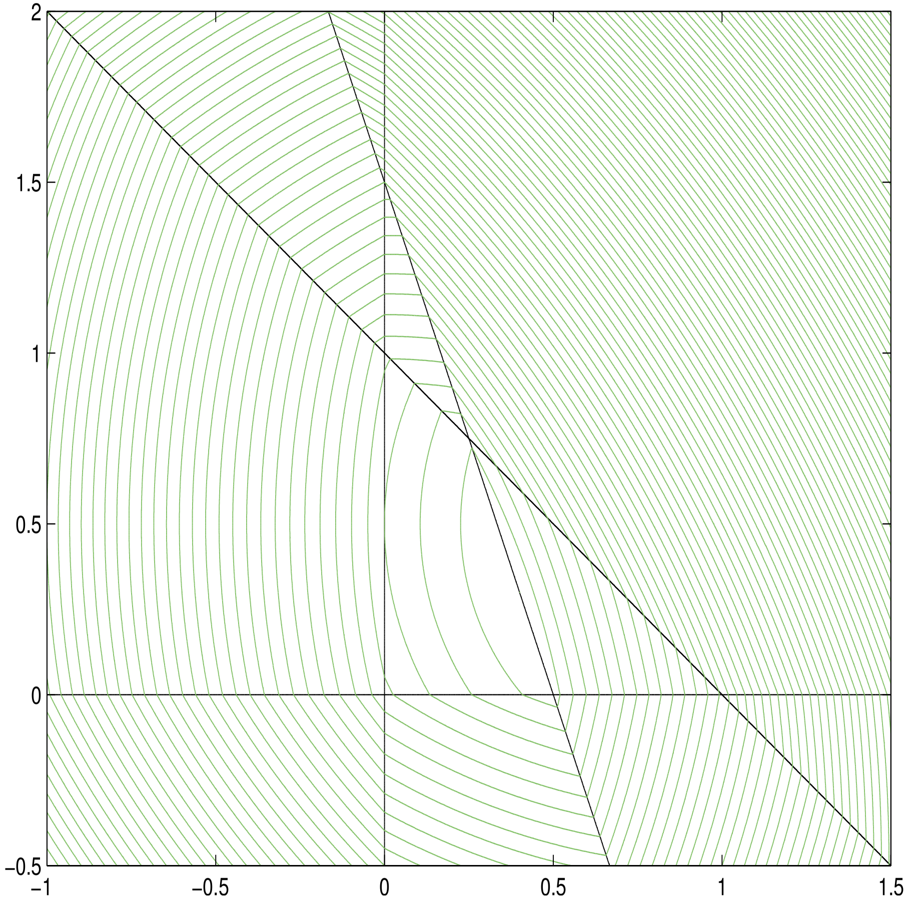
- ⊙ Big-M: for some sufficiently large M

$$\underset{x \in \mathbb{R}^n,\, \xi \in \mathbb{R}}{\text{minimize}} \;\; q(x) + M\xi \;\; \text{subject to} \;\; Ax + \xi r \geq b \;\; \text{and} \;\; \xi \geq 0$$

- ⊙ ℓ_1QP $(\rho > 0)$ — may be reformulated as a QP

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; q(x) + \rho \| \max(b - Ax, 0) \|$$

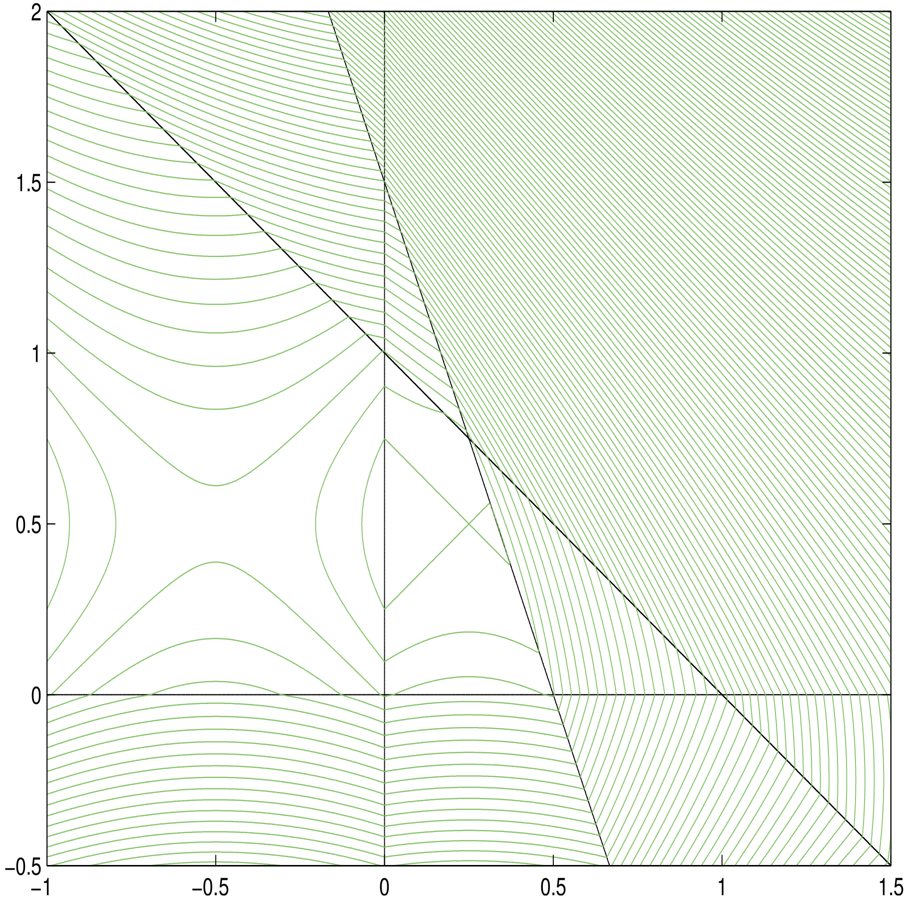
CONVEX EXAMPLE



$$\min (x_1 - 1)^2 + (x_2 - 0.5)^2$$
$$\text{subject to } \; x_1 + x_2 \leq 1$$
$$3x_1 + x_2 \leq 1.5$$
$$(x_1, x_2) \geq 0$$

Contours of penalty function $q(x) + \rho \| \max(b - Ax, 0) \|$ (with $\rho = 2$)

NON-CONVEX EXAMPLE



$$\min -2(x_1 - 0.25)^2 + 2(x_2 - 0.5)^2$$
$$\text{subject to } \; x_1 + x_2 \le 1$$
$$3x_1 + x_2 \le 1.5$$
$$(x_1, x_2) \ge 0$$

Contours of penalty function $q(x) + \rho \| \max(b - Ax, 0) \|$ (with $\rho = 3$)

TERMINATION, DEGENERACY & ANTI-CYCLING

So long as $\alpha_k > 0$, these methods are finite:

- finite number of steps to find an EQP with a feasible solution
- finite number of EQP with feasible solutions

If x_k is degenerate (active constraints are dependent) it is possible that $\alpha_k = 0$. If this happens infinitely often

- may make no progress (a cycle) $\Longrightarrow$ algorithm may stall

Various anti-cycling rules

- Wolfe's and lexicographic perturbations
- least-index — Bland's rule
- Fletcher's robust method

NON-CONVEXITY

- ⊙ causes little extra difficulty so long as suitable factorizations are possible
- ⊙ **Inertia-controlling** methods tolerate at most one negative eigenvalue in the reduced Hessian. Idea is

  1. start from working set on which problem is strictly convex (e.g., a vertex)
  2. if a negative eigenvalue appears, do not drop any further constraints until 1. is restored
  3. a direction of negative curvature is easy to obtain in 2.

- ⊙ latest methods are not inertia controlling $\Longrightarrow$ more flexible

COMPLEXITY

- When the problem is convex, there are algorithms that will solve QP in a polynomial number of iterations
  - some interior-point algorithms are polynomial
  - no known polynomial active-set algorithm
- When the problem is non-convex, it is unlikely that there are polynomial algorithms
  - problem is NP complete
  - even verifying that a proposed solution is locally optimal is NP hard

NON-QUADRATIC OBJECTIVE

When $f(x)$ is **non quadratic**

- ⊙ $H = H_k$ changes
- ⊙ active-set subproblem

$$x_{k+1} \approx \arg\min f(x) \quad \text{subject to} \quad a_i^T x = [b]_i \quad \text{for all} \quad i \in \mathcal{W}_k$$

  - ⋄ iteration now required but each step satisfies $A_k s = 0$
    $\Longrightarrow$ linear algebra as before
  - ⋄ usually solve subproblem inaccurately
    - ▹ when to stop?
    - ▹ which Lagrange multipliers in this case?
    - ▹ need to avoid zig-zagging in which working sets repeat