Optimization and Optimal Control

Engineers like to do things well

$\Rightarrow$ optimization.

Contents of this lecture:

- Unconstrained optimization, including algorithms:
  - steepest descent,
  - conjugate gradient,
  - Newton, pseudo-Newton
- Solution of nonlinear vector equations
- Generalized matrix inverses and least-squared error solution of $y = Ax$
- Duality and constrained optimization, (convex analysis used)
- Introduction to interior-point methods
- Introduction to dynamical optimization $\rightarrow$ optimal control theory

Some notation:

Suppose:

a statement $\alpha \Rightarrow \Box$ — a contradiction (eg. $(\beta=0)\wedge(\beta>0)$)

Conclusion: α must be false

Hence one can prove that θ is true by showing

$\neg\theta \Rightarrow \Box$ — a contradiction

not

called a proof by contradiction

A motivation for a basic optimization problem

$x \longrightarrow$ System or design $\longrightarrow V(x)$

x = vector of parameters

$V(x)$ = system's operating cost or cost of design

Could be a usual financial cost perhaps including penalties for not meeting specifications, etc.

A basic optimization problem

parameter vector $x \in \mathbb{R}^n$

cost function $V : \mathbb{R}^n \to \mathbb{R}$

$$\boxed{\begin{array}{l} \text{minimize } V(x) \\ x \in F \end{array}}$$

feasible set = the set of x's which satisfy the constraints

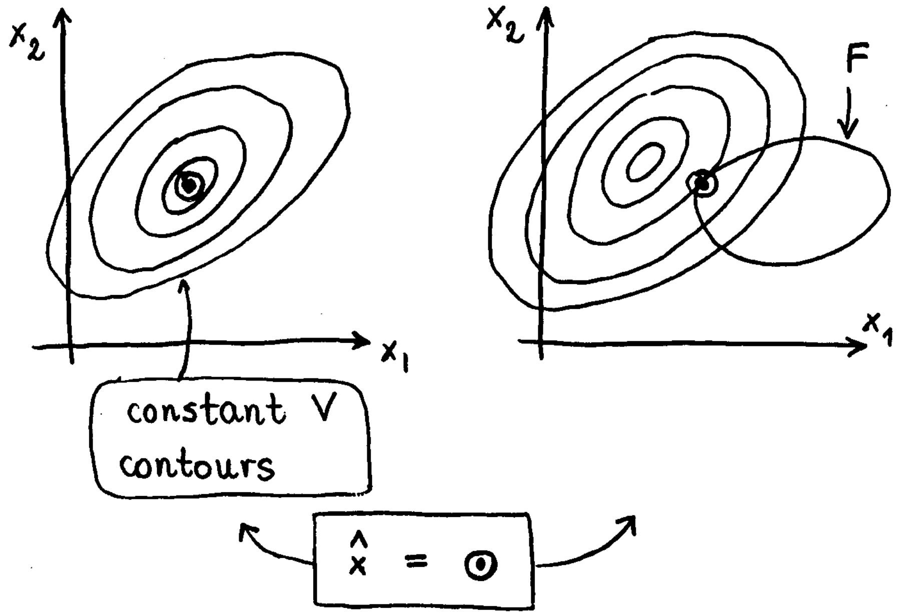
= eg. the set of allowable parameter vectors, bearing in mind constraints

e.g. temperature constraints for chemical processes, size and strength constraints for mechanical engineering designs, etc.

Examples : $n = 2$

unconstrained opt. $F = \mathbb{R}^2$

constrained opt. $F \subsetneq \mathbb{R}^2$



Meaning of optimization:

(1.1) $\hat{x}$ = a global minimizer of V on F

iff

$$[\hat{x} \in F] \wedge [V(x) \geq V(\hat{x}),\ \forall x \in F]$$

Set of all global minimizers :

$$\boxed{\arg\min_{x \in F} V(x)}$$

Optimal value :

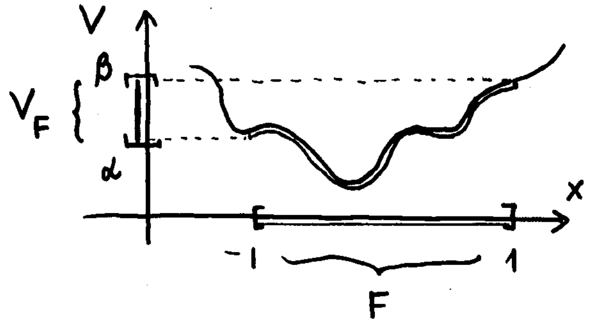
$$\boxed{\hat{V} \triangleq \min_{x \in F} V(x)}$$

Existence of an $\hat{x}$?

When might there not be a g.m. ?

$$\boxed{\begin{array}{l} (\alpha, \beta] \triangleq \{z \in \mathbb{R} : \alpha < z \leq \beta\} \\ [\alpha, \beta) \triangleq \{z \in \mathbb{R} : \alpha \leq z < \beta\} \end{array}}$$

Cost value set: $V_F \triangleq \{V(x) : x \in F\}$



(1.2) Th. (i) $\exists$ g.m. if eg. $V_F = [\alpha, \beta]$

(ii) $\nexists$ g.m. if eg. $V_F = (\alpha, \beta]$

Proof [of (ii)] : — $\boxed{V_F = (\alpha,\beta]}$

We will show that :

\# : $[\hat{x} = \text{a g.m.}] \Rightarrow \square$ — a contra-diction

therefore : $\nexists$ any g.m. — because $\hat{x}$ is just a name for any g.m.

Proof of \# :

Suppose $[\hat{x} = \text{a g.m.}]$

$\Downarrow$

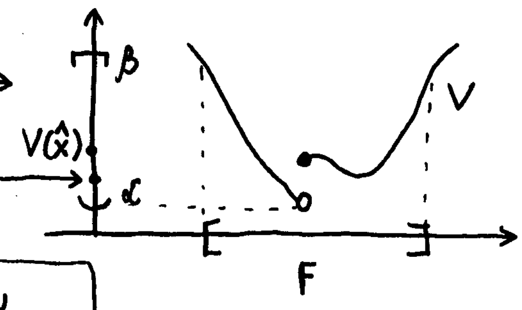
$[\hat{x} = \text{a g.m.}] \wedge [V(\hat{x}) \in V_F]$

$\Downarrow$

$\Downarrow$

We have this situation →

$\exists$ some θ →



because $V(\hat{x}) > \alpha$ as $V(\hat{x}) \in V_F = (\alpha, \beta]$

$\Downarrow$

$[\hat{x} = \text{a g.m.}] \wedge [\exists\, \theta \in V_F \text{ with } \theta < V(\hat{x})]$

$\Downarrow$ — because $V_F = \{V(x) : x \in F\}$

$[\hat{x} = \text{a g.m.}] \wedge [\exists\, \tilde{x} \in F \text{ with } V(\tilde{x}) = \theta < V(\hat{x})]$

$\Downarrow$ — because $\tilde{x}$ is feasible and gives lower V than $\hat{x}$ does, which contradicts the optimality of $\hat{x}$

$\square$ — a contra-diction

QED

Bounded sets

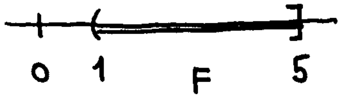
(1.3) $F \subset \mathbb{R}^n$ called bounded iff :

$\exists \ r \in [0,\infty)$ so that

$$\|x\| < r \ , \ \forall x \in F$$

e.g. $F \subset \mathbb{R}$

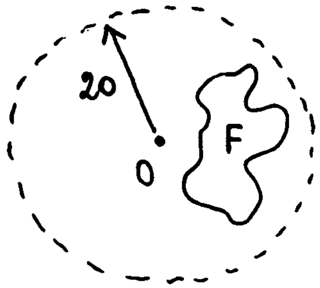
$\|x\| = |x|$
as x = scalar



bounded
e.g. use $r = 6$

e.g. $F \subset \mathbb{R}^2$

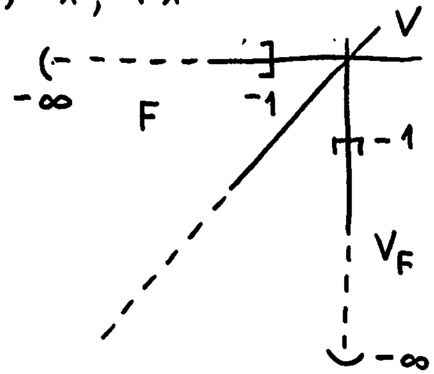
bounded
e.g. use $r = 20$



I Fact :

a g.m. $\hat{x}$ might not exist if F is NOT bounded

e.g. $\underset{\substack{\| \\ (-\infty,-1]}}{F} \subset \mathbb{R}$; $V(x) = x, \ \forall x$



$V_F = (-\infty, -1]$

$\Downarrow$

$\nexists$ a g.m.

because F is not bounded

by Th. (1.2) because $V_F = (\alpha, \beta]$ for $\alpha = -\infty, \beta = -1$

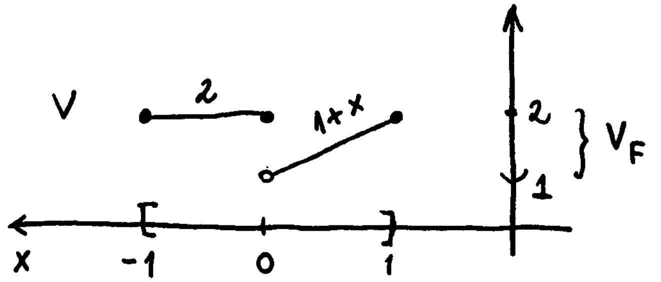
$-\infty$ has occured here because F is not bounded

II Fact: $\boxed{\text{a g.m. } \hat{x} \text{ might not exist if } V \neq \text{continuous on } F}$

$\boxed{V \text{ is continuous at every } x \in F}$

e.g. $F = [-1, 1] \subset \mathbb{R}$

$$V(x) = \begin{cases} 2, & x \leq 0 \\ 1+x, & x > 0 \end{cases}$$



$V_F = (1, 2]$

$\Downarrow$ $\boxed{\text{Th. (1.2)}}$

no g.m. $\boxed{\text{because of the discontinuity of } V}$

Limit points and closed sets

(1.4) $p \in \mathbb{R}^n$ is called a Lt. pt. of $F \subset \mathbb{R}^n$

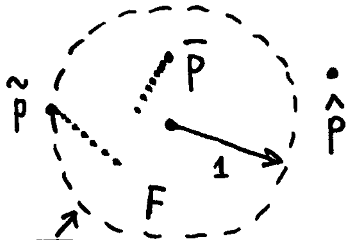
iff: there exist points $q \in F$, $q \neq p$ which are arbitrarily near p #

i.e. iff: you can get as close as you like to p while remaining in F (and while not being equal to p)

e.g. $F = \{x \in \mathbb{R}^2 : \|x\| < 1\}$

	$\in F$	Lt. pt.
$\bar{p}$	Y	Y
$\tilde{p}$	N	Y
$\hat{p}$	N	N

not closed



F is called closed if it contains all its limit points

(but $F = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$ <u>is</u> closed)

\# more precisely (for math. friends only):

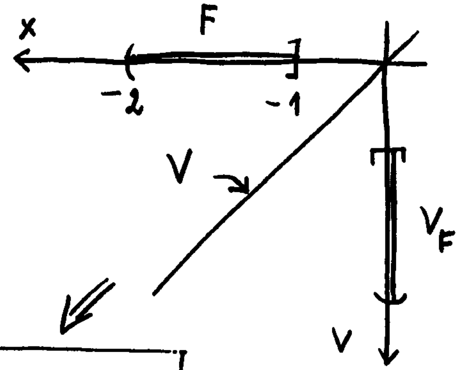
iff: for every $\varepsilon > 0$,

$\exists\, q \in F$, with $q \neq p$, such that $\|p - q\| < \varepsilon$

III Fact : there might not exist a g.m. $\hat{x}$ if $F \neq$ closed

e.g. $F = (-2,-1]$, $V(x) = x$, $\forall x$

not closed :
$-2 =$ lt. pt.
$-2 \notin F$



no g.m.
because V_F has the form $(\alpha,\beta]$, which has happened because $F \neq$ closed

Examples shown :

might not be a g.m. $\hat{x}$ if :

- $F \neq$ bounded
- $F \neq$ closed
- $V \neq$ continuous

Suggests :

(1.5) Th : $\exists$ a g.m. $\hat{x}$ if :

- $F =$ closed & bounded
- $V =$ continuous on F

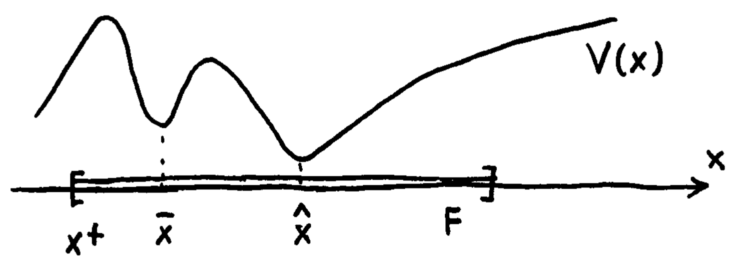
Proof : too hard

But : Existence of a g.m. matters as there is not any hope of using an algorithm to find a g.m. if one does not exist ! Also , this topic deepens understanding of what optimization is about.

Assumption :

For V & F in this course : $\exists$ a g.m. $\hat{x}$

Local minimizers



$\hat{x}$ = the g.m.

$\hat{x}, x^+, \bar{x}$ = local minimizers

(1.6) $\tilde{x}$ = a local minimizer of V on F

iff:

$$[\tilde{x} \in F] \wedge \left[V(x) \geq V(\tilde{x}), \ \forall \underline{\text{feasible } x} \atop \underline{\underline{\text{near } \tilde{x}}} \right]$$

$$\boxed{\begin{array}{l} \text{i.e. } x \text{ such that} \\ [x \in F] \wedge [\| x - \tilde{x} \| < \varepsilon] \\ \text{for suitably small } \varepsilon > 0 \end{array}}$$

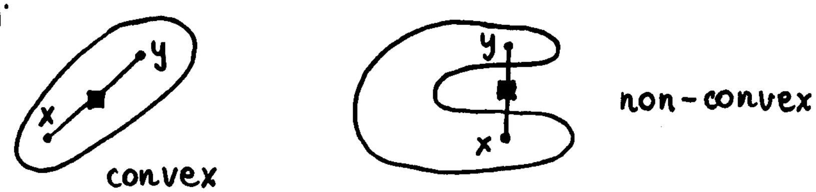
Convexity

... a useful property of many V & F.

A set <u>F is convex</u> iff

every point along the line joining any two points in F is itself in F

e.g.



<u>Algebraic characterization</u>:

(1.7) F is convex iff

$$\underbrace{\alpha x + (1-\alpha) y}_{\blacksquare} \in F$$

$\forall \alpha \in [0,1]$

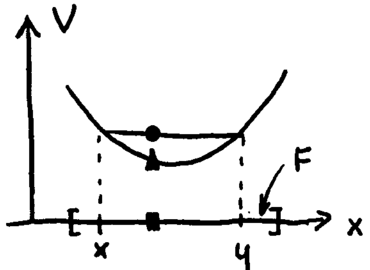
$\forall x, y \in F$

Convex functions

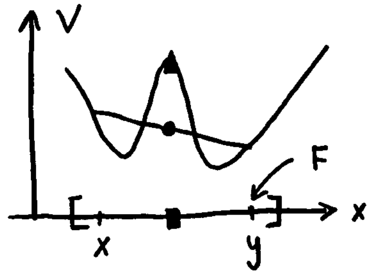
(1.8) A function $V : \mathbb{R}^n \to \mathbb{R}$ is called <u>convex on a convex set F</u> iff:

$$\underset{\blacktriangle}{V}[\underbrace{\alpha x + (1-\alpha) y}_{\blacksquare}] \leq \underbrace{\alpha V(x) + (1-\alpha) V(y)}_{\bullet}$$

$\forall\, \alpha \in [0,1]$, $\forall\, x, y \in F$



V = convex on F



V $\neq$ convex on F

$\Downarrow$

> So: for V to be convex on F : V must lie below (i.e. $\leq$) the straight line interpolating between $V(x)$ and $V(y)$ everywhere between x and y, and for all $x, y \in F$.

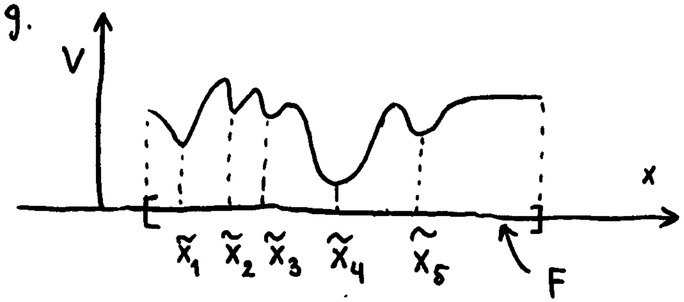
Significance of convexity

Most opt. algs. find only a local minimizer (l.m.)

In general : to find a g.m. :

- first find all l.m.s — generally impossible
- select the best

e.g.



$\hat{x} = \tilde{x}_4$

So, in general

finding a g.m. $\neq$ possible, BUT

$\Downarrow$

(1.9) **Th** Suppose :

V is convex on convex F

$\exists$ a g.m.

Then : all l.m.s of V on F are also g.m.s.

SO :

> For convex V & F :
> we can stop optimizing once any l.m. is found, as it must also be a g.m.

Proof : (by showing that any l.m. $\bar{x}$ is actually a g.m.)

(only for $n = 1$, but generalizes easily)

Say : $\hat{x}$ = a g.m.

$\bar{x}$ = a l.m. with $\hat{x} \neq \bar{x}$

Clearly : $V(\bar{x}) \geqslant V(\hat{x})$

but

(#) $[V(\bar{x}) > V(\hat{x})] \Rightarrow \square$ a contradiction

hence $V(\bar{x}) = V(\hat{x})$

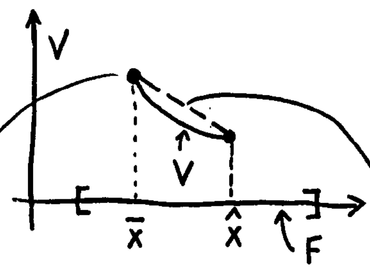
$\Rightarrow$ l.m. $\bar{x}$ has the same cost as g.m. $\hat{x}$

$\Rightarrow$ $\bar{x}$ = a g.m. too

QED.

<u>Proof of (#)</u> : $[V(\bar{x}) > V(\hat{x})]$

$\Downarrow$



$\Rightarrow \square$ a contradiction

contradicts $\bar{x}$ = a l.m.

this situation arises because, as V is a convex function on F, V lies below the line interpolating between $V(\bar{x})$ and $V(\hat{x})$.

QED

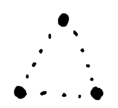
Algorithms for unconstrained optimization on $\mathbb{R}^n$

$F = \mathbb{R}^n$ — no constraints

Simplex algorithm

A simplex in $\mathbb{R}^n$ = set of $n+1$ equidistant points

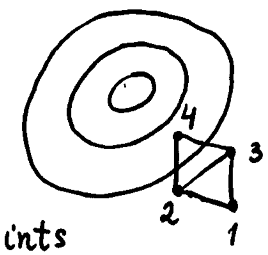
e.g. in $\mathbb{R}^2$:



Basic simplex alg.

— not the same as the Simplex Alg. of Linear Programming

- start with simplex (1,2,3)
- reflect the point in the simplex (1,2,3) which has greatest cost in the centroid of the remaining points giving a new simplex (2,3,4) "nearer" a l.m.

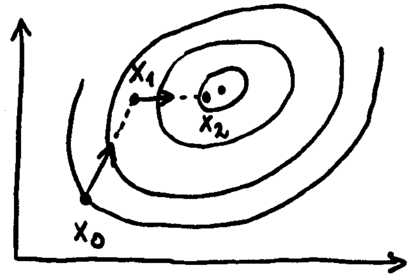


- repeat for simplex (2,3,4) etc.

Properties: simple in principle but needs modifications to help convergence to a l.m.

Search direction algorithm

$x_1 = x_0 + \omega_0 s_0$



- Choose $x_0 \in \mathbb{R}^n$ (an estimate of a g.m. $\hat{x}$)
- Set iteration index $j=0$

1) [Iteration j]:
   - Using info available at x_j, choose a search direction $s_j \in \mathbb{R}^n$ #
   - choose a scalar $\omega_j > 0$ so $V(x_j + \omega_j s_j) < V(x_j)$
   - Set $x_{j+1} := x_j + \omega_j s_j$, $j := j+1$ and go to 1)

choose s_j so V should decrease as one moves along s_j, away from x_j

Remarks:

very important type of alg. as it can be developed & adapted a lot

Problems:

choice of s_j, ω_j

Concerning choice of s_j:

The Gradient

$\mathbb{R}^2$ case

Under weak conditions ← $\boxed{\text{see cond. (a) \& (b) on p. 2.4}}$

$$V\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}\right] \overset{\Omega}{=} V\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] + \frac{\partial V}{\partial x_1}[x]\,\delta x_1 + \frac{\partial V}{\partial x_2}[x]\,\delta x_2$$

$$= V[x] + \underbrace{\begin{bmatrix} \partial V[x]/\partial x_1 \\ \partial V[x]/\partial x_2 \end{bmatrix}}_{\nabla V[x]}{}^T \underbrace{\begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}}_{\delta x}$$

$\boxed{\text{the gradient of } V \text{ at } x}$

i.e.

$$V[x+\delta x] \underset{\boxed{\text{for small } \delta x}}{\overset{\Omega}{=}} V[x] + \nabla V[x]^T \delta x$$

For $\mathbb{R}^n$ case:

(2.1) **Th** Consider $V: \mathbb{R}^n \to \mathbb{R}$ when the p.ds. $\frac{\partial V}{\partial x_i}[z]$:

(a) all exist $\forall z$ s.t. $\|z-x\| < \varepsilon$ for some $\varepsilon > 0$ #

(b) all are continuous at $z = x$

Then:

V is called differentiable at x and, for small δx:

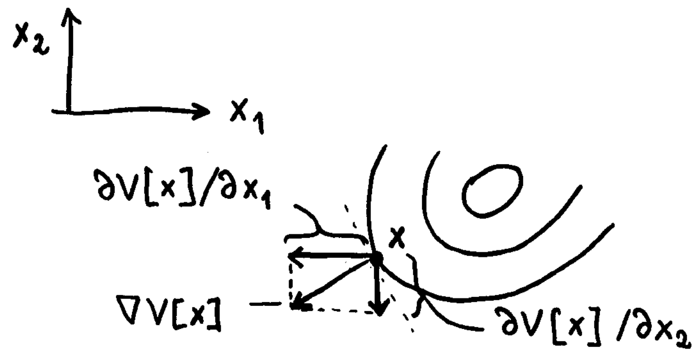
$$V[x+\delta x] \overset{\Omega}{=} V[x] + \underbrace{\nabla V[x]}{}^T \delta x$$

$$\nabla V[x] \triangleq \begin{bmatrix} \partial V[x]/\partial x_1 \\ \cdots\cdots \\ \partial V[x]/\partial x_n \end{bmatrix} \in \mathbb{R}^n$$

$\boxed{\text{the gradient of } V \text{ at } x}$

\# i.e. the p.ds. all exist everywhere near z

Graphical representation of ∇V



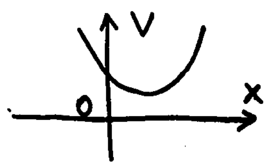
(2.2) Fact :

Even for the $\mathbb{R}^n$ case:

$\nabla V[x]$ is orthogonal to the constant V contour through x and points uphill

Quadratics on $\mathbb{R}^n$:

Quadratic of $x \in \mathbb{R}$



$$V(x) = \underset{\in \mathbb{R}}{a} + \underset{\in \mathbb{R}}{b}x + \frac{1}{2}x \underset{\in \mathbb{R}}{c} x, \quad V(x) \in \mathbb{R}$$

$\in \mathbb{R}$ & symmetric #

for a min : $c > 0$

in the sense that a scalar c can be thought of as a $[x]$ matrix, which is symmetric

Standard quadratic of $x \in \mathbb{R}^n$:

$$V(x) = a + b^T x + \frac{1}{2} x^T C x$$

$V(x) \in \mathbb{R}$, $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, $C \in \mathbb{R}^{n \times n}$

$C^T = C$ & $C > 0$

read as : C is positive definite & symmetric

$\equiv$ i.e $x^T C x > 0, \ \forall x \neq 0$, $x \in \mathbb{R}^n$

Tests for $C^T = C > 0$

For symmetric C:

$$[C > 0] \iff \left\{ \begin{array}{l} c_{11} > 0 \ ; \\ \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0 \ ; \\ \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} > 0 \ ; \\ \cdots\cdots \\ |C| > 0 \end{array} \right\}$$

Also:

$$[C>0] \Longleftrightarrow \lambda_i[C]>0, \ \forall i$$

the eigenvalues of C; since $C^T = C$, all the eigenvalues of C are real. This says that for C to be positive definite, all its eigenvalues must be strictly positive

By simple algebra:

$$V[x+\delta x] = V[x] + [b+Cx]^T\delta x + \underbrace{\tfrac{1}{2}\delta x^T C\,\delta x}_{\boxed{\text{second order}}} \quad (2.4)$$

$$\underset{\boxed{\text{for small } \delta x}}{\overset{\Omega}{=}} V[x] + [b+Cx]^T\delta x$$

Suggests #: $\nabla V[x] = b + Cx$ — check by partial differentiation

because we also have (from Th.(2.1)):

$$V(x+\delta x) \cong V(x) + \nabla V(x)^T\delta x$$

Hence:

(2.5) Th: For a standard quadratic on $\mathbb{R}^n$

(i) V is differentiable on $\mathbb{R}^n$
(i.e. V is differentiable at x, $\forall x \in \mathbb{R}^n$)

(ii) $\nabla V[x] = b + Cx$

(iii) $V[x+\delta x] = V[x] + \nabla V[x]^T\delta x + \frac{1}{2}\delta x^T C\,\delta x$

An optimality condition for general V

(2.6) Th If $V: \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\hat{x}$:

$$[\underbrace{\hat{x} = \text{a l.m. of } V \text{ on } \mathbb{R}^n}_{\boxed{\text{unconstrained}}}] \Rightarrow [\underbrace{\nabla V(\hat{x}) = 0}_{\boxed{\text{necessary condition for } \hat{x} = \text{a l.m.}}}]$$

Remark: So, one way to find a l.m. is:
solve $\nabla V(\hat{x}) = 0$ for $\hat{x}$
check if $\hat{x}$ = a l.m.

Proof of Th. (2.6):

$\boxed{\text{contrapositive i.e.:}}$

$$\neg\,[\nabla V(\hat{x})=0] \Rightarrow \neg\,[\hat{x} = \text{a l.m.}] :$$

$$\neg\,[\nabla V(\hat{x})=0] \Rightarrow [\nabla V(\hat{x}) \neq 0]$$

$\Rightarrow$ for all sufficiently small real $\omega > 0$

(i) $\hat{x} + \omega\,[-\nabla V(\hat{x})] = \text{near } \hat{x}$

(ii) $V(\hat{x} + \underbrace{\omega[-\nabla V(\hat{x})]}_{\boxed{\delta x}}) =$

$$\overset{\#}{\cong} V(\hat{x}) + \nabla V(\hat{x})^T \delta x$$

$$= V(\hat{x}) - \omega\,\|\nabla V(\hat{x})\|^2 < V(\hat{x})$$

$$\Rightarrow [\hat{x} \neq \text{a l.m.}] \Rightarrow \neg\,[\hat{x} = \text{l.m.}]$$

QED

\# the main point here is that the fact that

$$V(\hat{x}) + \nabla V(\hat{x})^T \delta x < V(\hat{x})$$

guarantees that

$$V(\hat{x} + \delta x) < V(\hat{x})$$

Application to minimization of a standard quadratic

(I) Solve $\nabla V(\hat{x}) = 0$, where $\nabla V(\hat{x}) = b + C\hat{x}$ $\boxed{\text{Th. (2.5 (ii))}}$

$\boxed{\text{Fact: } (C^T = C > 0) \Rightarrow C^{-1} \text{ exists}}$

Hence: $\boxed{\hat{x} = -C^{-1}b}$

(II) Check: $\hat{x} = \text{a l.m.}$?

$$V(x) = V(\hat{x} + \underbrace{[x - \hat{x}]}_{\delta x}) \overset{\boxed{\text{Th. 2.5 (iii)}}}{=}$$

$$= V(\hat{x}) + \underbrace{\nabla V(\hat{x})}_{=0}\delta x + \tfrac{1}{2}\underbrace{\delta x^T C \delta x}_{\geq 0 \text{ as } C > 0}$$

$$\geq V(\hat{x}), \quad \forall x \in \mathbb{R}^n$$

so $\hat{x}$ = a g.m. of V on $\mathbb{R}^n$ $\Downarrow$

(2.7) **Th**: $\hat{x} = -C^{-1}b$ is a unique g.m. of a quadratic V on $\mathbb{R}^n$ and:

$$V(\hat{x}) = a - \tfrac{1}{2} b^T C^{-1} b$$

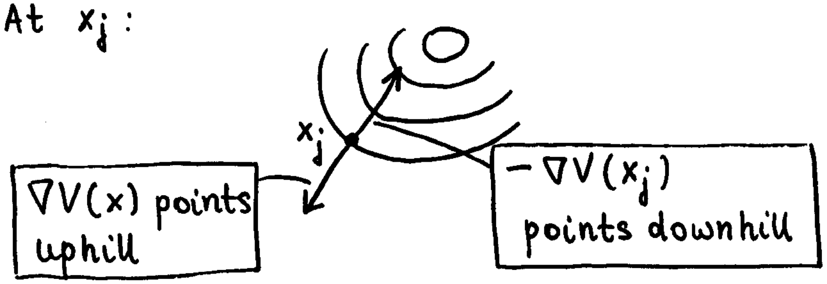
$\boxed{\text{not just a l.m.}}$

More about the Search Directions Alg.

A choice for s_j :

Assume that V is differentiable on $\mathbb{R}^n$

At x_j :



choose : $s_j = \underbrace{-\nabla V(x_j)}$

$\boxed{\text{steepest-descent search direction}}$

$\boxed{\text{as then } s_j \text{ points downhill so that } V \text{ should decrease as we move along } s_j \text{ from } x_j}$

The steepest-descent alg.

Choose $x_0 \in \mathbb{R}^n$. Set $j = 0$

1) [At iteration j]

- Stop if $\underbrace{\nabla V(x_j) = 0}$ — $\boxed{\text{as then } x_j \text{ satisfies a necessary condition for a l.m.}}$

  $\boxed{\text{in practice, use } \|\nabla V(x_j)\| < 10^{-8} \text{ say}}$

- Set : $s_j := -\nabla V(x_j)$
- Choose a scalar $\omega_j > 0$ so

  $V(x_j + \omega_j s_j) < V(x_j)$

  $\boxed{\text{often, } \omega_j \text{ chosen to minimize}_{\omega \in \mathbb{R}} \; V(x_j + \omega s_j) \text{ - at least approximately}}$

- Set : $x_{j+1} := x_j + \omega_j s_j$

  $j := j+1$

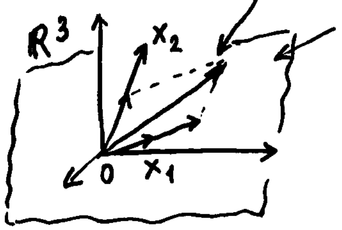
- Go to 1)

Some revision of linear algebra:

Consider $x^1, x^2, \dots, x^k \in \mathbb{R}^n$

A linear combination of these is $\sum_{i=1}^{k} \underbrace{\alpha_i}_{\boxed{\text{scalars}}} x^i$

e.g. $2x^1 + \frac{1}{2}x^2$



plane through $0, x^1, x^2$

$=$ set of all linear combinations of x^1, x^2

$= \{\alpha_1 x^1 + \alpha_2 x^2 : \alpha_1, \alpha_2 \in \mathbb{R}\}$

the set of vectors $\alpha_1 x^1 + \alpha_2 x^2$ that one gets as α_1, α_2 vary all over the set of all real numbers

$\triangleq$

$\mathcal{L}[x^1, x^2]$

the linear subspace spanned by x^1, x^2

We can say: $x^1, \dots, x^k \in \mathbb{R}^n$ are linearly independent iff:

$$\{0\} \subsetneqq \mathcal{L}[x^1] \subsetneqq \mathcal{L}[x^1, x^2] \subsetneqq \cdots \subsetneqq \mathcal{L}[x^1, \dots, x^k]$$

Then we say:

$\mathcal{L}[x^1]$ has dimension 1

$\mathcal{L}[x^1, x^2]$ –"– –"– 2

$\mathcal{L}[x^1, x^2, x^3]$ –"– 3 etc.

.....

Further: $x^1, \dots, x^n \in \mathbb{R}^n$ are linearly indep. iff $\mathcal{L}[x^1, \dots, x^n] = \mathbb{R}^n$

In general:

$\mathcal{L}[x^1, \dots, x^k]$

$\equiv$

the linear subspace spanned by $x^1, \dots, x^k$

$\equiv$

$\{\sum_{i=1}^{k} \alpha_i x^i : \alpha_i \in \mathbb{R}, \forall i\} \subset \mathbb{R}^n$

$\equiv$

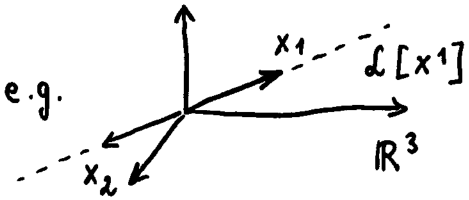
the plane through $0, x^1, \dots, x^k \subset \mathbb{R}^n$

Further: $x^1, \dots, x^k \in \mathbb{R}^n$ are linearly indep. iff:

$$\sum_{i=1}^{k} \alpha_i x^i = 0 \Rightarrow \alpha_i = 0 \ \forall i$$

Sometimes

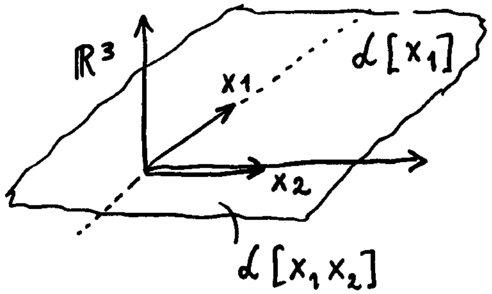
$\mathcal{L}[x^1] = \mathcal{L}[x^1, x^2]$ e.g.



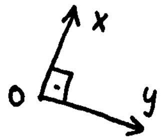
But usually

$\mathcal{L}[x^1] \subsetneqq \mathcal{L}[x^1, x^2]$

e.g.



Vectors $x, y \in \mathbb{R}^n$ are orthogonal



iff $x^T y = 0$

Notation:

s d alg. = search direction alg.
SD alg. = steepest-descent alg.

Properties of SD alg. for standard quadratic V on $\mathbb{R}^n$

$$\boxed{V(x) = a + b^T x + \tfrac{1}{2} x^T C x \quad , \; C^T = C > 0}$$

Why consider a quadratic ?

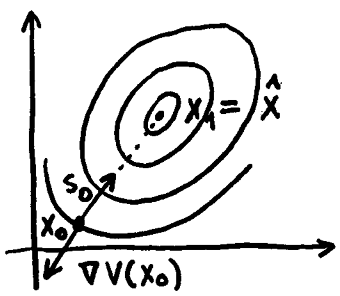
Often study algorithm behaviour on quadratics because :

- analysis easy
- [alg. good for quadratic V] $\Rightarrow$ [might be good for general V]
- [alg. bad for quadratic V] $\Rightarrow$ [alg. probably generally useless]

But, in practice, for a quadratic we would not usually find $\hat{x}$ using a s.d. alg. as $\hat{x} = -C^{-1}b$ for a quadratic

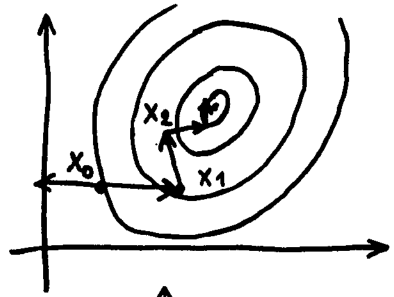
– so just solve $C\hat{x} = -b$

Characteristic of S.D. alg. with exact minimization along each s_j



can give $\hat{x}$ in one iteration

OR



can zig-zag giving slow descent

depending on x_0

(3.1) **Th** For • SD alg. with exact minimization along each s_j

• standard quadratic V on $\mathbb{R}^n$

(i) $[\nabla V(x_0) = \text{eigenvector of } C] \Rightarrow [x_1 = \hat{x}]$ – NICE

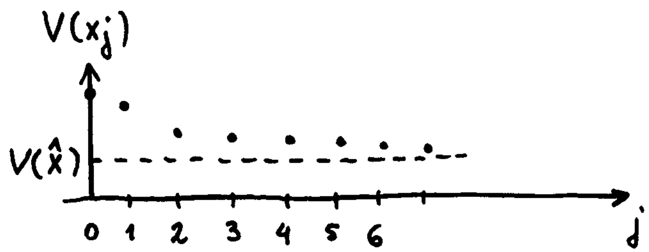
#

\# e is an eigenvector of C iff $\exists$ an eigenvalue λ_i of C and

$$(C - \lambda_i I)e = 0$$

(3.1) <u>Th</u> continued

(ii) $\left[\nabla V(x_0) \neq \text{eigenvector of } C\right] \Rightarrow \left[\begin{array}{l} x_j \neq \hat{x}, \\ \forall \text{ finite } j \end{array}\right]$ – BAD



(iii) $0 \leq [V(x_j) - V(\hat{x})] \leq \alpha^j [V(x_0) - V(\hat{x})]$

$$\boxed{0 \leq \alpha = 1 - \frac{\lambda_{min}(C)}{\lambda_{max}(C)} < 1} \quad \$$$

$\Downarrow$

(iv) $V(x_j) \longrightarrow V(\hat{x})$

$\boxed{\text{smaller } \alpha \text{ guarantees faster convergence}}$

\$ Here $\lambda_{min}(C)$ is the smallest eigenvalue of C and $\lambda_{max}(C)$ is the largest

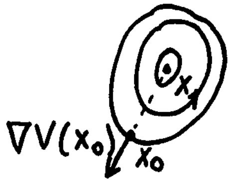
<u>Justification of (3.1) Th. (i)</u>:

$\boxed{\text{no proof for (ii) \& (iii) here}}$

for a quadratic $\hat{x} = -C^{-1}b$

so, for $x_1 = \hat{x} = -C^{-1}b$

needed that:



$x_1 - x_0 \parallel \nabla V(x_0)$

$\boxed{\text{parallel}}$

$\Downarrow$ $\exists\, \lambda_i \neq 0$ s.t. $\nabla V(x_0) + \lambda_i (x_1 - x_0) = 0$

$$\equiv \nabla V(x_0) + \lambda_1(-C^{-1}b - x_0) = 0$$

$$\equiv \nabla V(x_0) - \lambda_1 C^{-1}\underbrace{(Cx_0 + b)}_{\boxed{\nabla V(x_0)}} = 0$$

$\Downarrow$

$$C\,\nabla V(x_0) - \lambda_1 \nabla V(x_0) = 0$$

$\boxed{\nabla V(x_0) \text{ is an eigenvector of } C}$

QED

<u>Remark</u>: In view of the possibility of the zig-zag effect, SD is not a very good alg. BUT it is simple, robust

$\Rightarrow$ can be useful

$\Rightarrow$ an improved alg. for optimizing quadratics needed

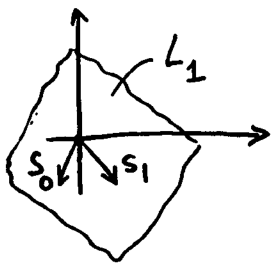
First study :

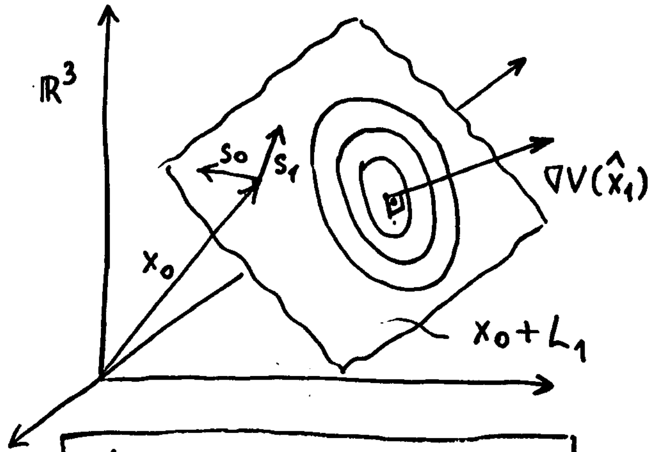
<u>Optimization of a quadratic on a linear variety</u>

$s_0, s_1, s_2, \dots \in \mathbb{R}^n \longleftarrow$ $\boxed{\text{search directions}}$

$\underbrace{\text{L.S.S.}}_{\boxed{\text{linear subspace}}}$: $L_j := \mathcal{L}[s_0, s_1, \dots, s_j]$

<u>Linear variety</u> : $x_0 + L_j := \{x_0 + y : y \in L_j\}$





unique for our standard quadratic

$\hat{x}_j$ denotes the global minimizer of V on variety $x_0 + L_j$

Suggests :

(3.2) $\underline{\underline{\text{Th}}}$: $\hat{x}_j$ minimizes V globally on $x_0 + L_j$ iff :

(i) $\hat{x}_j \in x_0 + L_j$

(ii) $\nabla V(\hat{x}_j)^T s_i = 0$, $i = 0, 1, \dots, j$

$\boxed{\text{(proof only "geometric")}}$

<u>Conjugate direction algorithm for standard quadratic on $\mathbb{R}^n$</u>

CD alg = sd alg. when :

(i) exact minimization used along each s_j

(ii) s_i are C-conjugate, in that

$$s_i^T C s_j = 0 \quad , \quad i \neq j \qquad (\pounds)$$

(3.3) $\underline{\underline{\text{Th.}}}$: For a CD algorithm :

x_j minimizes V globally on any variety $x_0 + L_{j-1}$, $j = 1, 2, \dots, n$

<u>Proof:</u> By Th (3.2) we just have to show that:

(a) $x_j \in x_0 + L_{j-1}$

(b) $\nabla V(x_j)^T s_i = 0 \quad , \quad i = 0, 1, \ldots, j-1$

<u>Proof of (a):</u> $(x_j \in x_0 + L_{j-1})$

$$x_1 = x_0 + \underbrace{w_0 s_0}_{\in \mathcal{L}[s_0] = L_0} \in x_0 + L_0$$

$$x_2 = x_1 + w_1 s_1 = x_0 + \underbrace{w_0 s_0 + w_1 s_1}_{\in \mathcal{L}[s_0 s_1] = L_1} \in x_0 + L_1$$

$$x_3 = x_2 + w_2 s_2 = x_0 + \underbrace{w_0 s_0 + w_1 s_1 + w_2 s_2}_{\in L[s_0, s_1, s_2] = L_2} \in x_0 + L_2$$

$\cdots \cdots$

$$\Downarrow$$

$$x_j \in x_0 + L_{j-1}$$

QED

<u>Proof of (b):</u> $(\nabla V(x_j)^T s_i = 0,\ i = 0, 1, \ldots, j-1)$

$$\equiv$$

$(\nabla V(x_j)^T s_i = 0,\ i = 1, \ldots, j-1)$

$$\nabla V(x) \underset{\text{Th (2.5)}}{=} b + Cx$$

$$\nabla V(x + \delta x) = b + C(x + \delta x)$$
$$= \underbrace{b + Cx}_{\nabla V(x)} + C\delta x$$

(3.4) $$\boxed{\nabla V(x + \delta x) = \nabla V(x) + C\delta x \quad , \forall x, \delta x}$$

Also if V is minimized exactly along each s_j then:

$$\boxed{\omega_j = -\frac{\nabla V(x_j)^T s_j}{s_j^T C s_j} \quad , \forall j}$$

this is because

$$\omega_j = \underset{\omega \in \mathbb{R}}{\arg\min}\ V(x_j + \omega s_j) \quad , \forall j$$

$$\Downarrow$$

ω_j must solve $\frac{d}{d\omega} V(x_j + \omega s_j)_{|\omega = \omega_j} = 0$

$$\Downarrow$$

$$\nabla V(\underbrace{x_j + \omega_j s_j}_{\shortparallel})^T \cdot s_j = 0 \quad , \forall j$$

$$s_j^T [b + C(x_j + s_j \omega_j)] = 0 \quad , \forall j$$

$$\equiv$$

$$\equiv$$

$$\underbrace{s_j^T b + s_j^T C x_j}_{s_j^T \nabla V(x_j)} + s_j^T C s_j \omega_j = 0 \ , \ \forall i$$

$$\equiv$$

$$\boxed{s_j^T \nabla V(x_j) + s_j^T C s_j \omega_j = 0 \ , \ \forall i \ \Rightarrow (3.4)}$$

So:

($) $\nabla V(x_1)^T s_0 = \nabla V(x_0 + \omega_0 s_0)^T s_0$

$$\underset{(3.4)}{=} \nabla V(x_0) + \omega_0 s_0^T C s_0 = 0$$

$$\boxed{\text{as } \omega_0 = -\frac{\nabla V(x_0)^T s_0}{s_0^T C s_0}}$$

We will show next that, because the s_i are C-conjugate, the fact that

$\nabla V(x_1)^T s_0 = 0$

causes that

$\nabla V(x_i)^T s_0 = 0$ for all $i > 1$.

$$\underbrace{\nabla V(x_2)^T s_0}_{\substack{\| \\ \nabla V(x_1 + \omega_1 s_1) \\ \| \\ \nabla V(x_1) + \omega_1 C s_1}} = \underbrace{\nabla V(x_1)^T s_0}_{\overset{(\$)}{=} 0} + \omega_1 \underbrace{s_1^T C s_0}_{\overset{(£)}{=} 0} = 0 \qquad (*)$$

$$\underbrace{\nabla V(x_3)^T s_0}_{\substack{\| \\ \nabla V(x_2 + \omega_2 s_2) \\ \| \\ \nabla V(x_2) + \omega_2 C s_2}} = \underbrace{\nabla V(x_2)^T s_0}_{\overset{(*)}{=} 0} + \omega_2 \underbrace{s_2^T C s_0}_{\overset{(£)}{=} 0} = 0$$

. . . .

So

$\nabla V(x_1)^T s_0 = \nabla V(x_2)^T s_0 = \dots = \nabla V(x_j)^T s_0 \overset{(\$)}{=} 0$

Similarly

$\nabla V(x_2)^T s_1 = \dots = \nabla V(x_j)^T s_1 = 0$

$\nabla V(x_3)^T s_2 = \dots = \nabla V(x_j)^T s_2 = 0$

. . . . . .

$\nabla V(x_j)^T s_{j-1} = 0$

QED
for Th (3.3) (b)

(3.5) Corollary For the CD alg.:

$$\underbrace{V(x_0) \geq V(x_1) \geq \cdots \geq V(x_n)}_{\text{as CD is a sd. alg.}} = V(\hat{x})$$

special property of CD alg.

Actually, CD might give $V(x_j) = V(\hat{x})$ for $j < n$ but this shows that at most n iterations are needed to achieve $V(\hat{x})$.

very NICE

especially as SD usually never achieves $V(\hat{x})$ exactly.

Proof of Corollary: $(V(x_n) = V(\hat{x}))$

Th (3.3) $\Rrightarrow$

x_n minimizes V on $x_0 + L_{n-1}$

$$L_{n-1} = \mathcal{L}[s_0, \ldots, s_{n-1}] \subset \mathbb{R}^n$$

But $\mathcal{L}[s_0, \ldots, s_{n-1}] = \mathbb{R}^n$

because:

Fact: C-conjugate $s_i \Rightarrow$ linear independence of all s_i, $i = 0, \ldots, n-1$

for if not, then: $s_j^T C s_i = 0 \quad \forall\, i \neq j$

and $\sum_{i=1}^{n} \alpha_i s_i = 0$ for some α_i s.t. $\sum_{i=1}^{n} |\alpha_i| \neq 0$

$\Downarrow$

$\sum_{i=1}^{n} \alpha_i C s_i = 0$ -"- -"- -"-

$\Downarrow$

$s_j^T \cdot \Big| \sum_{i=1}^{n} \alpha_i C s_i = 0$ -"- -"- -"-

$\Downarrow$

$\alpha_j \underbrace{s_j^T C s_j}_{>0 \text{ since } C > 0} = 0$ -"- -"-, $\forall j$

$\Downarrow$

$\alpha_j = 0 \quad \forall j \quad \Rightarrow \Box$ — a contradiction.

$\Downarrow$

$\Rightarrow$ s_i, $i = 0, \ldots, n-1$ must be linearly independent

So: x_n minimizes V on $x_0 + \mathbb{R}^n = \mathbb{R}^n$

$\Rightarrow V(x_n) = V(\hat{x})$

One way to generate C-conjugate s_i:

(3.6) <u>The Conjugate-Gradient Alg.</u>

Choose $x_0 \in \mathbb{R}^n$

set $s_0 = -\nabla V(x_0)$, $j=0$

1) Stop if $\| \nabla V(x_j) \| < \varepsilon \longleftarrow$ as then $x_j \cong \hat{x}$

Choose $\omega_j \in \arg\min_{\omega \in \mathbb{R}} V(x_j + \omega s_j)$

Set $x_{j+1} = x_j + \omega_j s_j$

$$\beta_{j+1} := \frac{[\nabla V(x_{j+1}) - \nabla V(x_j)]^T \nabla V(x_{j+1})}{\| \nabla V(x_j) \|^2} \in \mathbb{R}$$

$s_{j+1} := -\nabla V(x_{j+1}) + \underbrace{\beta_{j+1} s_j}$

$j := j+1$

Go to 1)

this term makes CG different from SD

(3.7) <u>Th</u> For CG alg. and quadratic V:

(i) the $s_0, s_1, \dots$ generated before it stops are C-conjugate

not proved here but easy to check eg. for s_0 and s_1:

$$\beta_1 = \frac{[\overbrace{\nabla V(x_1) - \nabla V(x_0)}^{= \omega_0 C s_0}]^T \nabla V(x_1)}{\underbrace{\| \nabla V(x_0) \|^2}_{= \|s_0\|^2}}$$

$\Downarrow$

$$s_0^T C s_1 = -s_0^T C \nabla V(x_1) + \underbrace{\omega_0}_{= \frac{\|s_0\|^2}{s_0^T C s_0} = \frac{-\nabla V(x_0)^T s_0}{s_0^T C s_0}} s_0^T C \frac{s_0^T C \nabla V(x_1)}{\|s_0\|^2} s_0 = 0$$

$\Downarrow$

(ii) CG alg. = a CD alg.

$\Downarrow \longleftarrow$ Th (3.3)

(iii) x_j minimizes V on $x_0 + L[s_0, \dots, s_{j-1}]$ for each x_j found before it stops

$\Downarrow$

(iv) CG alg. stops after $k \leq n$ iterations with $x_k = \hat{x}$ ← NICE

Comparison of CG and SD:

(3.8) Th If, for a quadratic V, SD and CG start at the same x_0:

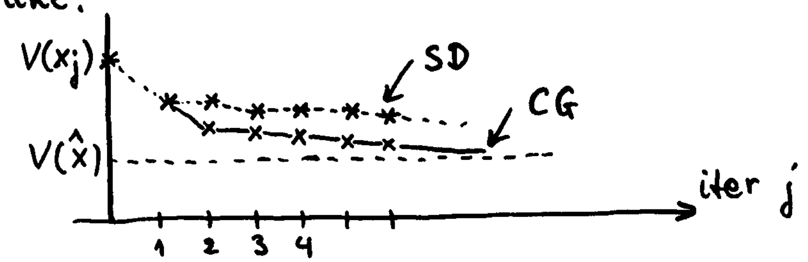
(a) $\underset{CG}{x_1} = \underset{SD}{x_1}$

(b) either $\underset{CG\ SD}{x_1} = \hat{x}$

or $\underset{CG}{V(x_j)} < \underset{SD}{V(x_j)}, \quad \forall j > 1$

hence CG has a theoretical advantage over SD

Remark: What usually happens is something like:



Newton's method

For suitably differentiable V:

$$V(x_j+\delta x) \cong V(x_j) + \nabla V(x_j)^T\delta x + \sum_{p=1}^{n}\sum_{q=1}^{n} \frac{\partial^2 V(x)}{\partial x_p \partial x_q}\delta x_p \delta x_q$$

$$= V(x_j) + \nabla V(x_j)^T \delta x + \frac{1}{2}\delta x^T V_{xx}(x_j)\delta x$$

where

(i) $V_{xx}(x) \in \mathbb{R}^{n\times n}$

(ii) $[V_{xx}(x)]_{p,q} \triangleq \dfrac{\partial^2 V(x)}{\partial x_p \partial x_q}$

(iii) $V_{xx}(x)$ is symmetric since

$$[V_{xx}(x)]_{p,q} = \frac{\partial^2 V(x)}{\partial x_p \partial x_q} = \frac{\partial^2 V(x)}{\partial x_q \partial x_p} = [V_{xx}(x)]_{q,p}$$

eg. $V: \mathbb{R}^2 \to \mathbb{R}$ then

$$V_{xx}(x) = \begin{bmatrix} \dfrac{\partial^2 V(x)}{\partial x_1 \partial x_1} & \dfrac{\partial^2 V(x)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 V(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 V(x)}{\partial x_2 \partial x_2} \end{bmatrix}$$

So:

$$V(x_j+\delta x) \underbrace{\hat{=}}_{\boxed{\text{for small } \|\delta x\|}} V(x_j) + \nabla V(x_j)^T\delta x + \frac{1}{2}\delta x^T V_{xx}(x_j)\delta x \qquad (4.1)$$

(4.2) $\boxed{\begin{array}{l}\text{For our standard quadratic on } \mathbb{R}^n \\ V(x_j+\delta x) = V(x_j) + \nabla V(x_j)^T\delta x + \frac{1}{2}\delta x^T C\,\delta x \\ \qquad\qquad \underline{\text{for all } \delta x \in \mathbb{R}^n} \\ \Rightarrow V_{xx}(x_j) = C > 0, \ \forall x_j \end{array}}$

<u>How to get search directions from such quadratic approximations of V ?</u>

Consider a slightly more general quadratic approximations

$$V(x) = V(x_j + \underbrace{[x - x_j]}_{\hat{=}\,\delta x})$$

$$\hat{=} V(x_j) + \nabla V(x_j)^T\delta x + \frac{1}{2}\delta x^T P(x_j)\delta x \triangleq v_j^P(x)$$

with $P(x_j)^T = P(x_j)$

$\boxed{\text{Standing for: value at } x \text{ of expansion about } x_j \text{ with second order contribution } \tfrac{1}{2}\delta x^T P(x_j)\delta x}$

(4.4) Th Consider

$$V_j^P(x) = V(x_j) + \nabla V(x_j)^T (x - x_j) + \tfrac{1}{2}(x - x_j)^T P(x_j)(x - x_j)$$

where $P(x_j)^T = P(x_j)$

Then:

$$\min_{x \in \mathbb{R}^n} V_j^P(x)$$

(I) exists and occurs uniquely at:

$$\tilde{x}_j = x_j - P(x_j)^{-1} \nabla V(x_j) \quad \text{if} \quad P(x_j) > 0$$

(II) exists and occurs non-uniquely at:

$$\tilde{x}_j = x_j - P(x_j)^{+} \nabla V(x_j)$$

$P(x_j)^{+}$: the so called pseudo-inverse of $P(x_j)$; to be defined later

if $P(x_j) \geqslant 0$ and $\nabla V(x_j) \in \mathcal{R}[P(x_j)]$

$P(x_j) \geqslant 0$: i.e. $x^T P(x_j) x \geqslant 0$ for all $x \in \mathbb{R}^n$, i.e. $\lambda_i[P(x_j)] \geqslant 0, \forall i$

$\mathcal{R}[P(x_j)]$: range of $P(x_j)$ $\equiv \{x \in \mathbb{R}^n : P(x_j)z = x, z \in \mathbb{R}^n\}$

(III) does not exist if any $\lambda_i[P(x_j)] < 0$

e.g. to see (I):

$$\nabla V_j^P(\tilde{x}_j) = \nabla V(x_j) + P(x_j)(\tilde{x}_j - x_j) \underset{!}{=} 0$$

necessary condition for a l.m. of V_j^P

$$\Rightarrow \tilde{x}_j - x_j = -P(x_j)^{-1} \nabla V(x_j)$$

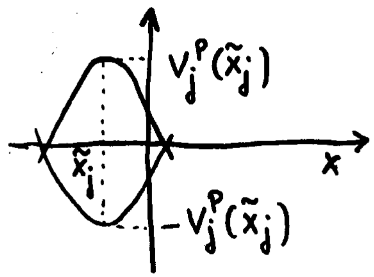
to see III:

suppose then that e.g. $P(x_j) < 0$

$\Rightarrow$ any $\tilde{x}_j$ s.t. $\nabla V_j^P(\tilde{x}_j) = 0$ is a maximizer (NOT a minimizer) of V_j^P since

$$\max_{x \in \mathbb{R}^n} V_j^P(x) = -\min_{x \in \mathbb{R}^n} \left[-V_j^P(x)\right]$$

minimizer exists and is given by (I) since $-V_j^P$ involves $-P(x_j) > 0$

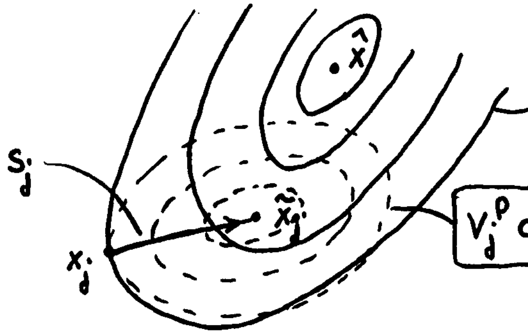


Hence minimization of V_j^P is simplest when $P(x_j) > 0$ and $P(x_j) = P(x_j)^T$. ← henceforth always assumed

<u>For general V</u> :

$$V(x) \cong V_j^P(x) \quad \text{when} \quad x \cong x_j$$

and we might have



Suggests : use $s_j = \tilde{x}_j - x_j$ ← g.m. for V_j^P

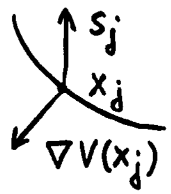
> Note: Better approximation V_j^P to V
>
> $\approx\!>$ $\tilde{x}_j$ nearer $\hat{x}$
>
> $\Rightarrow$ s_j better, because it points nearer to $\hat{x}$

(4.6) **Def^n** s_j is called a descent-direction if:

$$V(x_j + \omega s_j) < V(x_j)$$

for some $\omega > 0$

e.g. if $\nabla V(x_j)^T s_j < 0$ #



(4.7) **Th** If $P(x_j) > 0$ and $\nabla V(x_j) \neq 0$ and $s_j \triangleq \tilde{x}_j - x_j$

Then :

① $s_j = -P(x_j)^{-1} \nabla V(x_j)$ ← see Th. (4.4)(I)

② $s_j =$ a descent direction

to see this :

$$\nabla V(x_j)^T s_j = -\nabla V(x_j)^T P(x_j)^{-1} \nabla V(x_j) < 0$$

because $P(x_j) > 0 \Rightarrow P^{-1}(x_j) > 0$ for all x_j ($\not{S}$)

\# as then $\nabla V(x_j)$ and s_j are at an obtuse angle to each other

Proof of ($):

call $y \triangleq P^{-1}(x_j)x \Rightarrow x = P(x_j)y\ ,\ \forall x$

$\Rightarrow x^T P^{-1}(x_j)x = y^T P(x_j)\underbrace{P^{-1}(x_j)P(x_j)}_{\boxed{=I}}y$

$= y^T P(x_j) y > 0\ ,\ \forall y \neq 0$

$\Downarrow$

$x^T P^{-1}(x_j)x > 0\ ,\ \forall x \neq 0$

$\boxed{\text{as } x = 0 \Leftrightarrow y = 0}$

Hence we consider:

The P-algorithm = the s.d. algorithm with the above s_j, $\forall j$

$\boxed{\text{different } P\text{'s yield different algorithms}}$

Some choices for $P(x_j)$

Requirements : $P(x_j)^T = P(x_j) > 0\ ,\ \forall x_j$

$V(x) \cong V_j^P(x)$ (*)

$\boxed{\text{near as possible}}$

$\Downarrow$

(I) If $V_{xx}(x_j) > 0\ ,\ \forall j$:

choose $P(x_j) = V_{xx}(x_j)$ #

$\Rightarrow s_j = -V_{xx}(x_j)^{-1}\nabla V(x_j)$

$\boxed{\text{Newton s.d. \& resulting P-alg. called Newton alg.}}$

(II) If $V_{xx}(x_j) \not> 0$:

choose $P(x_j)$ = a positive definite approximation to $V_{xx}(x_j)$

$\Downarrow$

modified Newton algorithm

Choosing a positive definite approximation to $V_{xx}(x_j)$: $\Downarrow$

As then $P^T = P > 0$ and $V_j^P(x)$ is a good approximation to $V(x)$ when $\|x - x_j\|$ is small — so the requirements (*) on P are satisfied

Put $V_{xx}(x_j)$ in spectral form:

$$V_{xx}(x_j) = M \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} M^T$$

$\perp$

orthogonal matrix (indicated by $\perp$) meaning that $M^T M = M M^T = I$ (the columns of M are normalized eigenvectors of V)

Facts: (1) $V_{xx} \not> 0$ iff any $\lambda_i \leq 0$

(2) $P > 0$ iff all $\lambda_i[P] > 0$

$\Downarrow$

Obtain a pos. def. approximation P to V_{xx} by setting

$$\bar{\lambda}_i = \lambda_i \quad \text{if } \lambda_i > 0$$
$$\bar{\lambda}_i = \varepsilon > 0 \quad \text{if } \lambda_i \leq 0$$

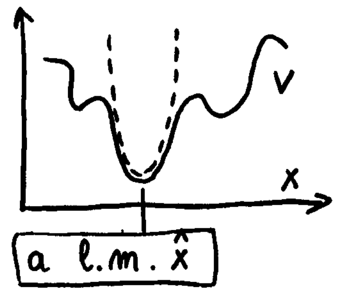
$\Downarrow$

$$P(x_j) = M \begin{bmatrix} \bar{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \bar{\lambda}_n \end{bmatrix} M^T \leftarrow \boxed{\text{pos. def. because all the } \bar{\lambda}_i > 0}$$

Remarks on Newton and modified-Newton algs.

(1) For our standard quadratic, Newton gives $x_1 = \hat{x}$, $\forall x_0$ $\leftarrow$ $\boxed{\text{very good}}$

(2) Many nasty functions V look like a quadratic near a l.m.



Hence Newton often gives fast convergence to a l.m. $\hat{x}$ once it gets near $\hat{x}$

(3) Say $n = 20$. Allowing for symmetry V_{xx} involves 210 different p.ds. $\frac{\partial^2 V}{\partial x_p \partial x_q}$

$\Rightarrow$ you have to find 210 different formulae

– too much like hard work

$\Rightarrow$ methods using V_{xx} not often practical for big problems $\leftarrow$ Bad

$\Downarrow$

Secant algorithms invented.

Secant Algorithms

In these estimate $V_{xx}(x_j)^{-1}$ from changes in ∇V and x observed as the sd. algorithm iterates

secant algorithms also known as quasi-Newton algorithms

Development of Secant Alg.

Consider the standard quadratic on $\mathbb{R}^n$ when:

- V & ∇V computable, somehow, $\forall x$
- C unknown (but $C^T = C > 0$)

corresponding to an unknown V_{xx} as $V_{xx}(x) = C$, $\forall x$, for quadratic V

Estimation of $V_{xx}(x)^{-1}$ = estimation of C^{-1}

For quadratic V:

$$\nabla V(x+\Delta x) = \nabla V(x) + C\,\Delta x \quad , \forall \Delta x$$

$$\Rightarrow \quad C\,\Delta x = \nabla V(x+\Delta x) - \nabla V(x) \triangleq \Delta g$$

$$\Downarrow$$

$$\Delta x = C^{-1}\Delta g \qquad (4.11)$$

given observed pairs $(\Delta x, \Delta g)$ shall use to estimate C^{-1}

note that $\Delta x = C^{-1}\Delta g$ for one pair $(\Delta x, \Delta g)$ does not specify every entry of C^{-1} — n linearly independent pairs $(\Delta x_i, \Delta g_i)$ are actually needed for that

$\Downarrow$ see next

At iteration j of the s.d. alg., WE KNOW:

past changes Δx_i , caused changes Δg_i

$\boxed{i=0,\dots,j-1}$

$$\Delta x_i \triangleq x_{i+1} - x_i \qquad \Delta g_i \triangleq \nabla V(x_{i+1}) - \nabla V(x_i)$$

$\Downarrow$

We know C^{-1} satisfies:

$$\left.\begin{aligned} &\Delta x_i = C^{-1} \Delta g_i \quad , i = 0,1,\dots,j-1 \\ &\text{with } (C^{-1})^T = C^{-1} > 0 \end{aligned}\right\} (\#)$$

Suppose we choose $H_j \in \mathbb{R}^{n \times n}$, so that:

$$\left\{\begin{aligned} &\Delta x_i = H_j \Delta g_i \quad , \forall i = 0,1,\dots,j-1 \\ &\text{with } H_j^T = H_j > 0 \end{aligned}\right.$$

Secant condition

Then $H_j \cong$ the unknown C^{-1}

in some sense
because H_j satisfies
properties # of C^{-1}

Since $H_j \triangleq C^{-1}$

Hence use: $\underbrace{s_j = -H_j \nabla V(x_j)}_{\text{pseudo-Newton search direction}} \triangleq \underbrace{-C^{-1} \nabla V(x_j)}_{\text{Newton search dir.}}$

pseudo-Newton search direction

Newton search dir. ($)
- would yield $x_{j+1} = \hat{x}$ (#)
- but unusable as C unknown

After optimizing along s_j (approximately?) we will have the extra information that

$$\Delta x_j = x_{j+1} - x_j = \omega_j s_j$$

caused by the gradient change

$$\Delta g_j = \nabla V(x_{j+1}) - \nabla V(x_j)$$

and we can build this new information into a potentially better estimate H_{j+1} of C^{-1}

than H_j

($) because $C = V_{xx}$ for our quadratic

(#) for our quadratic V

(4.12) <u>Secant Algorithm</u>

- Choose (i) $x_0 \in \mathbb{R}^n$
  (ii) symmetric, pos. def. $H_0 \in \mathbb{R}^{n \times n}$

  estimate of unknown C^{-1}, use $H_0 = I$ if no better guess available

- Set $j := 0$

1) [At iteration j]:

- Set $s_j := -H_j \nabla V(x_j)$ — pseudo-Newton s. dir
- Choose $\omega_j \geq 0$ so

$$V(x_j + \omega_j s_j) < V(x_j)$$

ideally ω_j chosen to minimize this

- Set $x_{j+1} := x_j + \omega_j s_j$
- Stop if $\|\nabla V(x_{j+1})\| < \varepsilon$ = small
- $\Delta x_j := x_{j+1} - x_j$, $\Delta g_j := \nabla V(x_{j+1}) - \nabla V(x_j)$
- Choose symm. pos. def. $H_{j+1} \in \mathbb{R}^{n \times n}$ so $H_{j+1} \cong C^{-1}$ – improved
- Set $j := j+1$, Go to 1)

There exist many ways for choosing suitable H_j. Most famous is:

The <u>Davidon-Fletcher-Powell Alg. (DFP)</u>

= the Secant Alg. when:
exact opt. used along each s_j
H_i chosen using:

$$H_{j+1} = H_j + \frac{\Delta x_j (\Delta x_j)^T}{(\Delta x_j)^T (\Delta g_j)} - \frac{H_j \Delta g_j (H_j \Delta g_j)^T}{(\Delta g_j)^T H_j \Delta g_j} \tag{4.13}$$

If approximate minimization used along each s_j, we have to use other update schemes for finding H_{j+1} from $H_j, \Delta x_j, \Delta g_j$

e.g. – Symmetric Rank
– BFGS — probably the best
– Broyden

Ref: R. Fletcher: "Practical Methods of Optimization",

Properties of Secant Algs. for quadratics

(4.14) Th Consider

- Secant Alg. using exact minimization along each s_j
- standard quadratic V

Then:

1. the s_j generated are C-conjugate

$\Rightarrow$ Secant Alg. = a C.D. Alg. – NICE
$\hat{x}$ achieved after at most n iterations

2. if all n iterations needed to find $\hat{x}$, then $H_n = C^{-1}$

the estimate H_j of C^{-1} is eventually exact

(4.15) Th. For

- many Secant Algs. using exact minimization along each s_j and $H_0 = I$

including DFP

- standard quadratic V:

$$\text{If} \quad \underbrace{x_0}_{\text{Secant}} = \underbrace{x_0}_{\text{CG}}$$

$$\text{then} \quad x_j = x_j \quad \text{for all } j \geq 1$$

Question:

Usually $H_0 = I$ so Secant & C.G. yield, for quadratics, the same x_j's

BUT:

Secant = more complex alg. than CG.

So: what is the use of Secant algs?

Answer:

Secant tends to be better than CG on non-quadratic V, when algs. applied properly to such V

Application of Algs. designed for quadratics to general nonlinear V

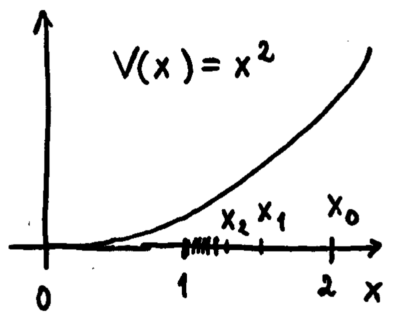
Even for nasty, differentiable, V :
s.d. algs. give

$$V(x_{j+1}) = V(x_j) \quad \text{if } x_j \neq \text{a l.m. } \hat{x}$$

Makes one think that #
always $x_j \xrightarrow[j\to\infty]{} [\text{a l.m. } \hat{x}]$

WRONG !

e.g. $F = \mathbb{R}$, $\hat{x} = 0$
<u>Alg</u> : $x_j = 1 + \frac{1}{j+1}$, $\forall j \geq 0$



$x_j \to 1 \neq \hat{x}$
even though
$V(x_{j+1}) < V(x_j)$, $\forall j$

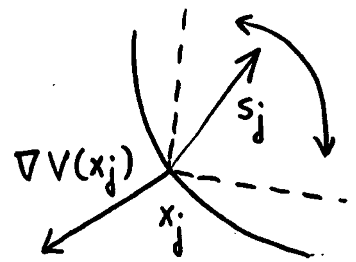
because the cost decreases at each iteration whenever $x_j \neq \hat{x}$

For a general function V :
- impossible to guarantee convergence of x_j to a l. m. $\hat{x}$

BUT
- convergence to a l. m. $\hat{x}$ is encouraged by using , $\forall j$:

(1) w_j = a fairly good approximation to $\hat{w}_j$ which minimizes V exactly along s_j

(2) an s_j satisfying an angle condition with respect to $\nabla V(x_j)$, i.e. a cone-condition :



(1) Determination of suitable ω_j

- For standard quadratic V:

$$\hat{\omega}_j := \arg\min_{\omega \in \mathbb{R}^n} V(x_j + \omega s_j)$$

$$= -\frac{[\nabla V(x_j)^T s_j]}{s_j^T C s_j}$$

- For more general V:

  $\nexists$ nice formula for $\hat{\omega}_j$

One practical way to estimate $\hat{\omega}_j$ is the Armijo Alg.

First note that:

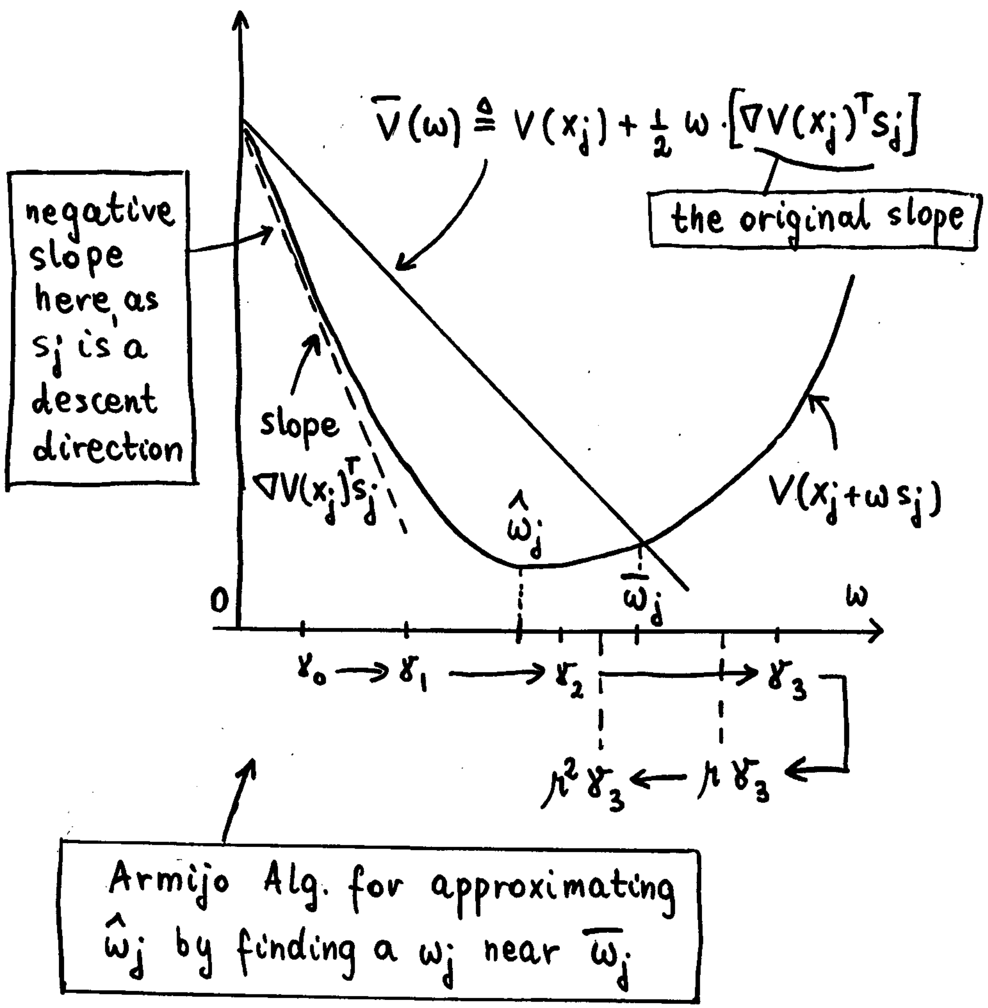
the slope of $V(x_j + \omega s_j)$ at $\omega = 0$ is:

$$\frac{d}{d\omega} V(x_j + \omega s_j)\Big|_{\omega=0} = \nabla V(x_j + \omega s_j)^T\Big|_{\omega=0} s_j$$

$$= \nabla V(x_j)^T s_j$$

A typical situation with s_j = a descent direction

– a typical plot of $V(x_j + \omega s_j)$ v.s. ω:



Armijo Alg.

ⓐ **Find a ω on right of $\bar{\omega}_j$**

Choose $\gamma > 1$ ($) (e.g. $\gamma = 1.5$).
Consider $\omega = \gamma^p$.
Increase p from 0 in unit steps until

$V(x_j + \gamma^p s_j) \geq \bar{V}(\gamma^p)$

for the first time.
Then: $\bar{\omega}_j \leq \gamma^p$ [$p = 3$ in example]

ⓑ **Find a ω_j on left of $\bar{\omega}_j$, near $\bar{\omega}_j$**

Choose $\mu \in (0,1)$, e.g. $\mu = 0.8$ (¢)
Consider $\omega = \mu^q \gamma^p$
Increase q from 0 in unit steps until ...

($) choose γ reasonably big so that one finds a point on the right of $\bar{\omega}_j$ for a small p i.e. using little work

(¢) choose μ reasonably near 1, so $\bar{\omega}_j$ is approximated reasonably accurately.

... until

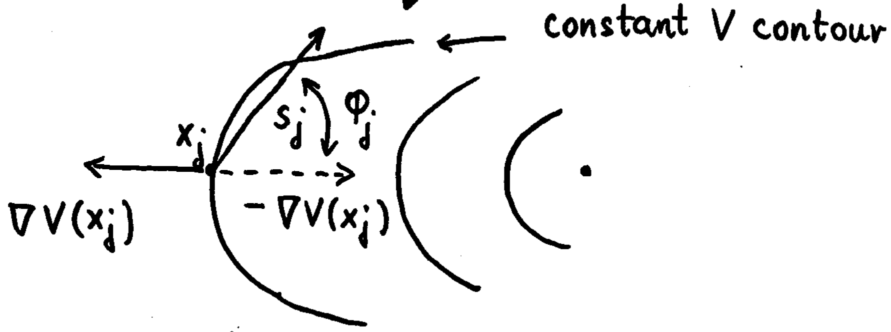
$V(x_j + \mu^q \gamma^p s_j) \leq \bar{V}(\mu^q \gamma^p)$

for the first time [$q = 2$ in example]

ⓒ Set $\underline{\omega_j = \mu^q \gamma^p}$

② **The cone condition**

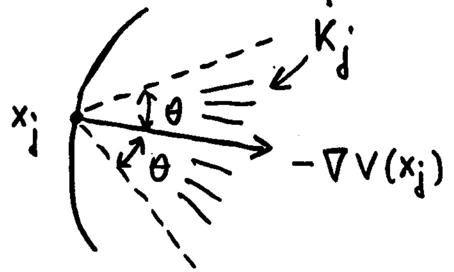
Not likely to get a big cost decrease along this s_j here



this is because s_j passes too near the contour (the tangent to the contour) through x_j,
i.e. because φ_j = too near 90°

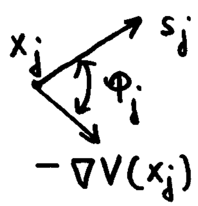
So: use a given s_j iff

$s_j \in \underbrace{\text{cone of acceptable search directions}}_{K_j}$



i.e.
$s_j \in K_j$ iff
$\varphi_j < \theta < 90°$
#

Test for $s_j \in K_j$:



I $\cos\varphi_j = \dfrac{s_j^T[-\nabla V(x_j)]}{\|s_j\| \, \|\nabla V(x_j)\|}$

II $\varphi_j < \theta$

iff $\cos\varphi_j > \cos\theta$

Hence : $s_j \in K_j$ iff :

$s_j^T[-\nabla V(x_j)] > \|s_j\| \, \|\nabla V(x_j)\| \cos\theta$

$\boxed{\text{easy to test}}$

\# which eliminates search directions passing too near the contour through x_j but not unduly restrictive

Action: If an algorithm finds an $s_j \notin K_j$

$\Rightarrow$ reset s_j to $-\nabla V(x_j) \in K_j$

$\boxed{\Rightarrow s_j \in K_j, \forall j \text{ – encouraging } x_j \to \text{l.m. } \hat{x}}$

C.G. for non-quadratic V

- C.G. designed for standard quadratics
  - for which $\hat{x} = -C^{-1}b$

$\Downarrow$

- C.G. not usually used for quadratics

BUT

- C.G. often works well for general V when applied sensibly

see next

\# Eventually this hope will be realized because functions V usually look like quadratics near a local minimizer

START C.G $\rightarrow x_0$

x_1

x_2

$\vdots$

x_n

C.G. stops at x_j if $\nabla V(x_j)=0$

If V acting like a quadratic : $\Rightarrow$ $\nabla V(x_j)=0$ for some $j \leq n$ $\leftarrow$ unlikely for $j<n$

If $\nabla V(x_n) \neq 0$, decide that V did not act like a quadratic, so RESTART CG. with new $x_0 =$ current x_n

AND HOPE # V will act like a quadratic this time

Also: at each j,

IF $s_j \notin K_j$

THEN: restart with new $x_0 =$ current x_j

$\Rightarrow$ s_j used $\in K_j$, $\forall j$

encourages $x_j \rightarrow$ l.m. $\hat{x}$

- -

For a general V:

quadratic theory NEVER exactly valid

$\Downarrow$

no need to use exact minimization along each s_j – which is required by CG. alg. for quadratics

$\Downarrow$

Use Armijo to approximate $\hat{\omega}_j$

---

because then the search direction used is the steepest descent search direction

Use of Secant Alg. with general V.

START with $H_0 = I$, x_0

If V acting like a quadratic: H_1, x_1

$V_{xx}(x_j) = C \;, \forall j$ $\quad H_2, x_2$

$\Downarrow$

$H_n = C^{-1}$

computed using update formula designed for quadratic V

$\Downarrow$ x_{n+1} should $= \hat{x}$

H_{n+1}, x_{n+1}

If $\nabla V(x_{n+1}) \neq 0$,
assume it is because V
has <u>not</u> so far been acting
like a quadratic $\rightarrow$ RESTART
with $H_0 = I$ ← (**)
new x_0 = current x_{n+1}

Also: if $s_j \notin K_j$ then RESTART
with $H_0 = I$
new x_0 = current x_{n+1}

(**) Discard H_{n+1} as it has been computed using an update formula designed for quadratics, and we have just decided that V has not been acting like a quadratic

Remark: <u>Basic ideas involved in applying Algs. designed for quadratics to general V</u>

(I) approximate $\hat{\omega}_j$ using Armijo Alg.

$\Rightarrow$ for Secant Alg. use
H_j update scheme valid for
approx. minimization along s_j

(II) restart (so as to use S.D. search dir.) whenever:

- alg. performance suggests V has not been acting like a quadratic
- $s_j \notin K_j$

Remark: Performance of algs. for general V

roughly:

- CG better than SD
  (as it is much better on quadratics)

- Secant algs. better than CG
  (as they approximate Newton)

- Newton – good in theory but impractical
  (as finding V_{xx} impractical)

- CG, Secant converge fast to l.m. $\hat{x}$ once they get near $\hat{x}$
  (as usually V looks like a quadratic near $\hat{x}$)

Remarks on calculating $\nabla V(x)$

Sometimes V is so complicated that finding formulae for the first-order p.ds. is too much like hard work

$\Rightarrow$ you may estimate each $\dfrac{\partial V(x)}{\partial x_i}$ using a finite difference method:

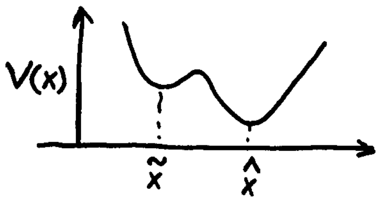
e.g. $$\frac{\partial V}{\partial x_1}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) \cong \frac{V\left(\begin{bmatrix} x_1+h \\ x_2 \end{bmatrix}\right) - V\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{h}$$

where h is chosen:

- small enough – so that a good approximation to the p.d. is obtained
- big enough – so that numerical noise does not give big errors

Choice of h is problem-dependent

(5.1) <u>A sufficient condition for unconstrained local optimality</u>



<u>Th (2.6) says:</u>

$$\left[\tilde{x} = \text{a local min'zer}\right] \Rightarrow \left[\nabla V(\tilde{x}) = 0\right]$$

$\boxed{\text{a necessary condition for local optimality}}$

"Stronger" result is:

(5.2) Th

$$\underbrace{\left\{\begin{matrix} \nabla V(\tilde{x}) = 0 \\ \text{and} \\ V_{xx}(\hat{x}) > 0 \end{matrix}\right\}}_{\boxed{\text{a sufficient condition for local optimality}}} \Rightarrow \left\{\begin{matrix} \tilde{x} \text{ is a local} \\ \text{minimizer of } V \end{matrix}\right\}$$

<u>Plausibilification</u>

For x near $\tilde{x}$:

$$\nabla V(x) \cong \nabla V(\tilde{x}) + \underbrace{\nabla V(\tilde{x})^T}_{0}(x - \tilde{x}) +$$

$$+ \frac{1}{2}(x-\tilde{x})^T \underbrace{V_{xx}(\tilde{x})}_{>0}(x-\tilde{x})$$

$$\geqslant V(\tilde{x}), \ \forall x \text{ near } \tilde{x}$$

i.e. $V(x) \geqslant V(\tilde{x}), \forall x$ near $\tilde{x}$

$\Rightarrow$ $\tilde{x}$ = a local minimizer of V on $\mathbb{R}^n$

<u>Example</u> $V : \mathbb{R}^2 \to \mathbb{R}$

$$V(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Might guess

$$\tilde{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ is a local minimizer}$$

Is it? Calculate:

$$\nabla V(\tilde{x}) = \begin{bmatrix} \partial V(\tilde{x})/\partial x_1 \\ \partial V(\tilde{x})/\partial x_2 \end{bmatrix}$$

$$\frac{\partial V}{\partial x_1}(x) = -400\,(x_2 - x_1^2)\,x_1 - 2(1 - x_1) = 0 \quad @\ \tilde{x}$$

$$\frac{\partial V}{\partial x_2}(x) = 200\,(x_2 - x_1^2) = 0 \quad @\ \tilde{x}$$

$$\nabla V(\tilde{x}) = 0$$

To check if $\tilde{x}$ = a loc. minimizer evaluate $V_{xx}(\tilde{x})$:

$$\frac{\partial^2 V(x)}{\partial x_1 \partial x_1} = -400\,x_2 + 1200\,x_1^2 + 2 = 802 \ @\ \tilde{x}$$

$$\frac{\partial^2 V(x)}{\partial x_1 \partial x_2} = \frac{\partial^2 V(x)}{\partial x_2 \partial x_1} = -400\,x_1 = -400 \ @\ \tilde{x}$$

$$\frac{\partial^2 V(x)}{\partial x_2^2} = 200.$$

So, $V_{xx}(\tilde{x}) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} > 0$

since $\det(802) = 802 > 0$

$$\det \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} > 0$$

Hence : $\tilde{x}$ is a local minimizer of V on $\mathbb{R}^n$

Revision of norms & orthogonal matrices

For $x \in \mathbb{R}^n$

$\|x\|$ = a measure of the "length of x"

many definitions possible – all satisfying :

(6.1) **Norm axioms** — chosen so $\|x\|$ behaves like length

$\forall\, x, z \in \mathbb{R}^n,\ \forall\, \alpha \in \mathbb{R}$:

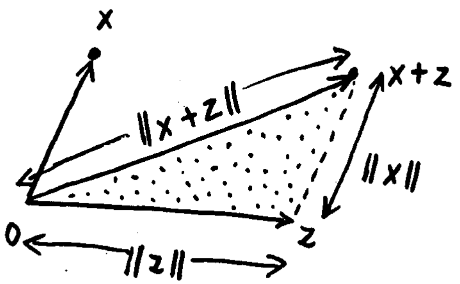
Axiom I : $\|x\| \geq 0$

Axiom II : $[\,\|x\| = 0\,] \Leftrightarrow [\,x = 0\,]$

Axiom III : $\|\alpha x\| = |\alpha| \cdot \|x\|$

Axiom IV : $\|x+z\| \leq \|x\| + \|z\|$

called triangle inequality -based on



Examples of norms on $\mathbb{R}^n$

1. $\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$

2. $\|x\|_2 \triangleq \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$ ← Euclidean norm

3. $\|x\|_\infty \triangleq \max\{|x_1|, |x_2|, \dots, |x_n|\}$

In this lecture always : $\|x\| = \|x\|_2$

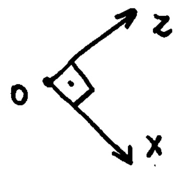
so : $\|x\|^2 = \sum_{i=1}^{n} x_i^2 = x_1^2 + \dots + x_n^2$

$$= (x_1, x_2, \dots, x_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x^T x$$

$$\Rightarrow \|x\| = (x^T x)^{1/2}$$

Orthogonal vectors :

In $\mathbb{R}^2$:



$x \perp z$

orthogonal

In $\mathbb{R}^n$?

(6.2) $[x \perp z] \underset{\text{by def.}}{\Longleftrightarrow} [\, \|x+z\| = \|x-z\| \,]$

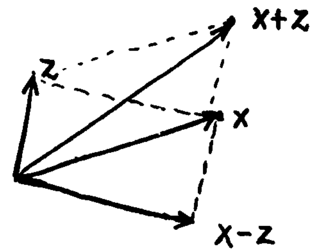
equivalently:

(6.3) $[x \perp z] \underset{\text{by def.}}{\Longleftrightarrow} [\, x^T z = 0 \,]$

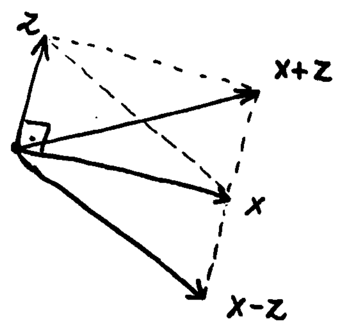
why this definition?

because it is a key property of orthogonality in $\mathbb{R}^2$

$x \not\perp z$



$\|x+z\| \neq \|x-z\|$



$\|x+z\| = \|x-z\|$

Equivalence of (6.2) & (6.3):

$$\|x+z\| = \|x-z\| \quad \leftarrow (6.2)$$

$$\Updownarrow$$

$$\|x+z\|^2 = \|x-z\|^2$$

$$\Updownarrow$$

$$(x+z)^T(x+z) = (x-z)^T(x-z)$$

$$x^Tx + \underbrace{z^Tx}_{x^Tz} + x^Tz + z^Tz = x^Tx - \underbrace{z^Tx}_{x^Tz} - x^Tz + z^Tz$$

$$\Updownarrow$$

$$2x^Tz = -2x^Tz$$

$$\Updownarrow$$

$$x^Tz = 0 \quad \leftarrow (6.3).$$

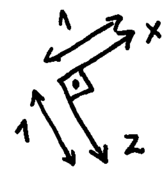
QED.

(6.4) Definitions

1. $x \in \mathbb{R}^n$ is called normalized iff $\|x\| = 1$
2. $x, z \in \mathbb{R}^n$ are called orthonormall iff:

(a) $x \perp z$

(b) $\|x\| = \|z\| = 1$



③ $x^1, x^2, \dots, x^k \in \mathbb{R}^n$ are called orthonormal iff:

(a) $x^i \perp x^j, \ \forall\, i,j \ (i \neq j)$

(b) $\|x^i\| = 1, \ \forall i$

e.g. the standard basis vectors for $\mathbb{R}^3$:

$$e_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad e_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad e_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

Orthogonal matrices (6.5)

very important for numerical algorithms

$P \in \mathbb{R}^{n \times n}$ is called orthogonal iff: its columns are orthonormal.

Some properties of $\perp P$: (6.6)

① $P^T P = P P^T = I \quad [\Rightarrow P^T = P^{-1}]$

② $\|Px\| = \|x\|$

because $\|Px\|^2 = (Px)^T Px = x^T \underbrace{P^T P}_{= I} x = x^T x = \|x\|^2$

The Cauchy-Schwartz inequality

For $x, y \in \mathbb{R}^n$:

(6.7) $|x^T y| \le \|x\|_2 \|y\|_2$

equality only when x & y point in the same direction

$\Rightarrow \quad x^T y \le \|x\|_2 \|y\|_2$

Proof: Obvious if $x = y = 0$, so, say $x \neq y$.

Minimize $\|x + \alpha y\|^2$ with respect to $\alpha \in \mathbb{R}$.

Since $\|x + \alpha y\|^2 = (x+\alpha y)^T (x + \alpha y)$

$= \|x\|^2 + 2\alpha x^T y + \alpha^2 \|y\|^2$

the minimizing α is:

$$\hat{\alpha} = -\frac{x^T y}{\|y\|^2}, \text{ and then}$$

$$0 \le \|x + \hat{\alpha} y\|^2 = \|x\|^2 - \frac{(x^T y)^2}{\|y\|^2}$$

so it is ≥ 0

$\Rightarrow (x^T y)^2 \le \|x\|^2 \|y\|^2$

$\Rightarrow |x^T y| \le \|x\| \|y\|$ QED

Linear least-squared error problems:

(formulation + background leading to Moore-Penrose pseudo-inverse)

Consider

$$\underbrace{y}_{\boxed{\text{output}}} = \underbrace{A}_{\boxed{\text{viewed as a transfer function of some non-dynamic system}}} \underbrace{x}_{\boxed{\text{input}}}, \quad A \in \mathbb{R}^{n\times n}$$

A control problem:

Given a desired output $y_d \in \mathbb{R}^n$

what makes $A x \underbrace{\cong}_{\boxed{\text{as near as possible}}} y_d$

If $\exists\, A^{-1} \Rightarrow$ required $x = A^{-1} y_d$

$$\Downarrow$$

$$\{ A x \underbrace{=}_{\boxed{\text{exactly}}} y_d \}$$

But what if $A \neq$ square? ($\Rightarrow$ when $\nexists\, A^{-1}$)?

In general:

suitable x can be found by solving:

$$\min_{x\in\mathbb{R}^n} \underbrace{\| y_d - Ax \|}_{\boxed{\text{distance of } Ax \text{ from } y_d}}$$

An obvious approach:

$$\min \| \cdots \| \iff \min \| \cdots \|^2$$

so: take $V(x) = \| y_d - Ax \|^2$

$$= (y_d - Ax)^T (y_d - Ax)$$

$$\Rightarrow \quad V(x) = \underbrace{\|y_d\|^2}_{a} + \underbrace{(-2A^T y_d)^T}_{b} x + \tfrac{1}{2} x^T \underbrace{[2A^TA]}_{C} x$$

But does $C^T = C$?

$$C^T = (2A^TA)^T \underbrace{=}_{\boxed{(AB)^T = B^TA^T}} 2A^T \underbrace{(A^T)^T}_{A} = C \quad \boxed{\text{YES}}$$

Is $C > 0$?

$$x^T C x = 2 \underbrace{x^T A^T}_{(Ax)^T} A x = 2 \|Ax\|^2 > 0, \ \forall x \neq 0$$

iff:

$Ax \neq 0, \forall x \neq 0$

$\Rightarrow$ $C > 0$ iff $Ax \neq 0$, $\forall x \neq 0$

So, suppose $Ax \neq 0$, $\forall x \neq 0$

Then $V(x) = \|y_d - Ax\|^2$

$\|$

a case of our standard quadratic

$$\Rightarrow \hat{x} = -C^{-1} b = \underbrace{(A^T A)^{-1} A^T y_d}_{\text{looks nice but ...}}$$

<u>Difficulty 1</u>: not always true that

$Ax \neq 0$, $\forall x \neq 0$

<u>Difficulty 2</u>: evaluating $A^T A$ very susceptible to errors caused by finite precision computing.

$\Rightarrow$ a better minimization method is desirable
- based on orthogonal matrices.

$\downarrow$

A geometrical interpretation

Set of all possible outputs $= \{Ax : x \in \mathbb{R}^n\}$

$\boxed{\text{general case} \\ A \text{ not square}}$

$\triangleq \mathcal{R}[A] \subset \mathbb{R}^m$

$\boxed{\text{the range of } A}$

<u>Computation of Ax</u>

we know: $(Ax)_i = \sum_{k=1}^{n} a_{ik} x_k$

Also:

column formula

$$Ax = \sum_{k=1}^{n} a_{*k} x_k$$

$\|$

$[a_{*1}, \dots, a_{*n}]$

column 1 of A

row formula

$$Ax = \begin{bmatrix} a_{1*} x \\ \vdots \\ a_{m*} x \end{bmatrix}$$

$\|$

$$\begin{bmatrix} a_{1*} \\ \vdots \\ a_{m*} \end{bmatrix} \leftarrow \text{row 1 of } A$$

So:

$$\mathcal{R}[A] = \{Ax : x \in \mathbb{R}^n\}$$

$$= \left\{ \sum_{k=1}^{n} a_{*k} x_k : x_k \in \mathbb{R}, \forall k \right\}$$

$\Rightarrow$ $\mathcal{R}[A]$ = set of all linear combinations of $a_{*1}, \dots, a_{*n}$

$= \mathcal{L}[a_{*1}, \dots, a_{*n}] \leftarrow$ the linear space spanned by the columns of A

Ideally would like an $x \in \mathbb{R}^n$ to exist such that $y_d = Ax$ which happens iff

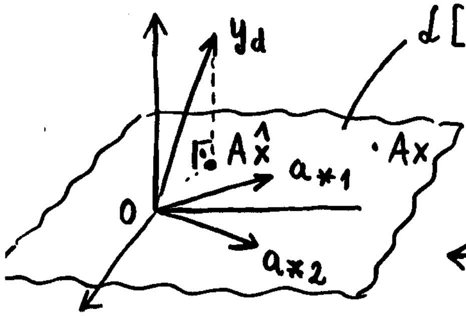
y_d = a possible system output, i.e. iff

$y_d \in \mathcal{R}[A]$ - set of all possible outputs

$\Downarrow$ then

$$\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2 = 0$$

<u>Example:</u> $(m = 3, n = 2)$



obvious geometrically that $y_d - A\hat{x}$ is $\perp$ to $\mathcal{R}[A]$, i.e is $\perp$ to any vector in $\mathcal{R}[A]$.

Actually, for a linear subspace like the range of A, the set of all vectors which are orthogonal to every vector in $\mathcal{R}[A]$ is called the <u>orthogonal complement</u> of the range of A, denoted $^{\perp}\mathcal{R}[A]$

Symbolically:

$$^{\perp}\mathcal{R}[A] = \{ z \in \mathbb{R}^n : z^T y = 0, \forall y \in \mathcal{R}[A] \}$$

Hence the fact that

$$(y_d - A\hat{x}) \perp \mathcal{R}[A]$$

can be expressed more neatly by:

$$(y_d - A\hat{x}) \in {}^{\perp}\mathcal{R}[A]$$

Algebraic proof that:

$\hat{x}$ minimizes $\|y_d - Ax\|^2$ on $\mathbb{R}^n$

if $(y_d - A\hat{x}) \perp \mathcal{R}[A]$ (even if $y_d \in \mathcal{R}[A]$)

$$\|y_d - Ax\|^2 = \|y_d - A(\hat{x} + x - \hat{x})\|^2$$

$$= \|(y_d - A\hat{x}) - A(x - \hat{x})\|^2$$

$$= \|y_d - A\hat{x}\|^2 - 2\underbrace{(y_d - A\hat{x})^T \underbrace{A\underbrace{(x-\hat{x})}_{\in \mathbb{R}^n}}_{\in \mathcal{R}[A]}}_{=0} + \underbrace{\|A(x-\hat{x})\|^2}_{\geq 0}$$

$$\geq \|y_d - A\hat{x}\|^2, \quad \forall x \in \mathbb{R}^n$$

$\Rightarrow$ $\hat{x}$ is a global minimizer if $(y_d - A\hat{x}) \perp \mathcal{R}[A]$.

QED

A^+ = the pseudo-inverse of A

called the Moore-Penrose generalized inverse

First introduce notation:

Partitioned vectors:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_m \end{pmatrix} \begin{matrix} \}\,\tilde{y} \\ \}\,\bar{y} \end{matrix} = \begin{pmatrix} \tilde{y} \\ \bar{y} \end{pmatrix}$$

$$\|y\|^2 = \sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{m} y_i^2 = \|\tilde{y}\|^2 + \|\bar{y}\|^2$$

Partitioned matrices / vectors:

$$Ax = \left[\begin{array}{c|c} B & C \\ \hline D & E \end{array}\right] \begin{pmatrix} \tilde{x} \\ \hline \bar{x} \end{pmatrix} = \begin{pmatrix} B\tilde{x} + C\bar{x} \\ \hline D\tilde{x} + E\bar{x} \end{pmatrix}$$

Derivation of A^+

$$\xrightarrow{x} \boxed{Ax = y} \longrightarrow y_d$$

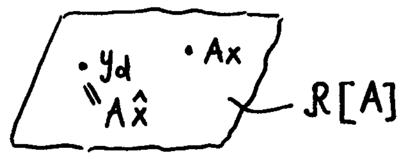
Ideally: choose x so $Ax = y_d$

possible if $y_d \in \mathcal{R}[A]$
impossible if $y_d \notin \mathcal{R}[A]$

Suggests:
choose x to min $\|y_d - Ax\|$

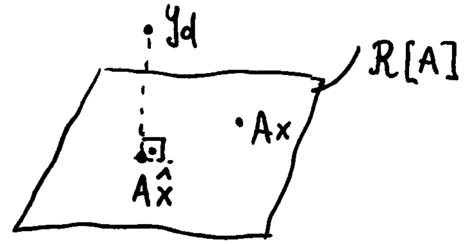
$\hat{x}$

If $y_d \in \mathcal{R}[A]$:



$\|y_d - A\hat{x}\| = 0$

If $y_d \notin \mathcal{R}[A]$:



$\|y_d - A\hat{x}\| > 0$

but as small as possible

Computation of an $\hat{x}$ aided by:

(7.0) Th For $0 \neq A \in \mathbb{R}^{m \times n}$, $\exists \perp P, \perp Q$ so:

$$A = P \left[\begin{array}{c|c} \hat{A}_{r \times r} & O_{r \times (n-r)} \\ \hline O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{array} \right] Q^T$$

where: $\hat{A}_{r \times r} \in \mathbb{R}^{r \times r}$ is invertible

$P \in \mathbb{R}^{m \times m}$

$Q \in \mathbb{R}^{n \times n}$

and $O_{m \times n}$ denotes an $m \times n$ matrix full of zeros

---

Also minimizing $\|y_d - Ax\|$

equivalent to

minimizing $\|y_d - Ax\|^2$

This is the general case: one might actually get:

$A = P \begin{bmatrix} \hat{A} \\ 0 \end{bmatrix} Q^T$, or $A = P[\hat{A} \; 0]Q^T$, or $A = P\hat{A}Q^T$

depending on A.

$$\boxed{\begin{array}{l} \text{Minimize } \|y_d - Ax\|^2 \\ x \in \mathbb{R}^n \end{array}}$$

$\|y_d - Ax\|^2$

$\| \quad \boxed{\|z\|^2 = \|P^T z\|^2 \text{ as } P \text{ is } \perp}$

$\|P^T [y_d - Ax]\|^2$

$\| \qquad P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} Q^T$

$$= \|\underbrace{P^T y_d}_{\begin{bmatrix} \tilde{y}_d \\ \bar{y}_d \end{bmatrix}} - \begin{bmatrix} \hat{A} & 0 \\ 0 & A \end{bmatrix} \underbrace{Q^T x}_{\begin{bmatrix} \widetilde{Q^T x} \\ \overline{Q^T x} \end{bmatrix}} \|^2$$

$\boxed{\widetilde{Q^T x} \text{ just denotes the top part of the partitioned vector } Q^T x \text{, etc.}}$

$$= \left\| \begin{bmatrix} \tilde{y}_d \\ \bar{y}_d \end{bmatrix} - \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{Q^T x} \\ \overline{Q^T x} \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} \tilde{y}_d - \hat{A}(\widetilde{Q^T x}) \\ \bar{y}_d \end{bmatrix} \right\|^2$$

$$= \underbrace{\|\tilde{y}_d - \hat{A}(\widetilde{Q^T x})\|^2}_{\boxed{\geq 0}} + \|\bar{y}_d\|^2 \geq \|\bar{y}_d\|^2, \forall x$$

$$= \|\bar{y}_d\|^2 \text{ whenever } \tilde{y}_d = \hat{A}(\widetilde{Q^T x})$$

Hence

$$\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2 = \| \bar{y}_d \|^2$$

$$\arg\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2 = \{ x \in \mathbb{R}^n : \tilde{y}_d = \hat{A}(\widetilde{Q^T x}) \}$$

$$= \{ x \in \mathbb{R}^n : (\widetilde{Q^T x}) = \hat{A}^{-1} \tilde{y}_d \}$$

$$= \left\{ x \in \mathbb{R}^n : Q^T x = \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix} \text{ for some } z \in \mathbb{R}^{n-r} \right\}$$

$\boxed{\text{i.e. } x = Q = \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix}}$

$$= \left\{ Q \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix} : z \in \mathbb{R}^{n-r} \right\}$$

which global minimizer is best ?
The global minimizers depend on z
Reasonable to use the shortest,
i.e. the shortest $Q \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix}$

Now:

$$\left\| Q \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ z \end{pmatrix} \right\|^2$$

since Q is orthogonal

$$= \| \hat{A}^{-1} \tilde{y}_d \|^2 + \| z \|^2$$

smallest for z=0

Hence use the global minimizer

$$\hat{x} = Q \begin{pmatrix} \hat{A}^{-1} \tilde{y}_d \\ 0 \end{pmatrix},$$

which is the x of least norm which minimizes $\| y_d - Ax \|^2$.

Then:

$$\hat{x} = Q \begin{pmatrix} \hat{A}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{y}_d \\ \bar{y}_d \end{pmatrix} = P^T y_d$$

$$= \underbrace{Q \begin{pmatrix} \hat{A}^{-1} & 0 \\ 0 & 0 \end{pmatrix} P^T}_{A^+} y_d$$

Called A pseudo-inverse

$\Rightarrow$ The general definition of A^+ is :

(7.1) For $A \in \mathbb{R}^{m \times n}$

$A^+ \in \mathbb{R}^{n \times m}$ is defined so $A^+ y_d$ is the x of least norm which minimizes :

$$\| y_d - Ax \|^2 \text{ on } \mathbb{R}^n$$

Aside

For $A \in \mathbb{R}^{m \times n}$

$A^r \in \mathbb{R}^{n \times m}$ is a right inverse of A if $A A^r = I$

$A^l \in \mathbb{R}^{n \times m}$ is a left inverse of A if $A^l A = I$

A has a normal inverse iff A has a left inverse and a right inverse (and both are equal)

A might have an A^l or an A^r even when A^{-1} does not exist !

Some properties of A^+

A	an $\hat{x}$ minimizing $\| y_d - Ax \|^2$
A^{-1} exists	$A^{-1} y_d$
A^{-1} does not exist	$A^+ y_d$

  $\Downarrow$

  if A has a (normal) inverse : $A^+ = A^{-1}$
- if A has a right-inverse : A^+ is one
- if A has a left-inverse : A^+ is one
- if $A = 0$: $A^+ = 0$

$\Downarrow$

A^+ = a fantastic , do-it-all inverse

Some applications of A^+

① **Choice of input x so $Ax = y_d$**

choose $x = A^+ y_d$ $\boxed{\Rightarrow Ax \text{ is as close as possible to } y_d}$

If A has a right-inverse, then:

$$Ax = \underbrace{AA^+}_{I} y_d = y_d \qquad \text{NICE}$$

$\boxed{\text{since } A^+ \text{ is a right-inverse if } A \text{ has one}}$

② **Recovery of x from the measurement $y = Ax$**

choose $\tilde{x} = A^+ y$ $\Rightarrow$ $\tilde{x}$ is a good estimate of the x generating y

If A has a left-inverse: then:

$$\tilde{x} = A^+ y = \underbrace{A^+ A}_{I} x = x$$

$\boxed{\text{since } A^+ \text{ is a left-inverse if } A \text{ has one}}$

i.e. A^+ recovers x without error

Computation of A^+

The detailed structure of the orthogonal decomposition of A affects formula + props. for A^+

A	A^+	Properties
$P\begin{bmatrix}\hat{A} & 0\\ 0 & 0\end{bmatrix}Q^T$	$Q\begin{bmatrix}\hat{A}^{-1} & 0\\ 0 & 0\end{bmatrix}P^T$	$A^+A \neq I$ i.e. $A^+ \neq$ a l-inv. $AA^+ \neq I$ i.e. $A^+ \neq$ a r-inv. $\nexists A^{-1}$
$P[\hat{A}\ 0]Q^T$	$Q\begin{bmatrix}\hat{A}^{-1}\\ 0\end{bmatrix}P^T$	$A^+A \neq I$ $AA^+ = I$ i.e. $A^+ =$ a r-inv. $\nexists A^{-1}$
$P\begin{bmatrix}\hat{A}\\ 0\end{bmatrix}Q^T$	$Q[\hat{A}^{-1}\ 0]P^T$	$A^+A = I$ i.e. $A^+ =$ a l-inv. $AA^+ \neq I$ $\nexists A^{-1}$
$P\hat{A}Q^T$	$Q\hat{A}^{-1}P^T$	$A^+ = A^{-1}$

Hence: calculation of A^+ easy once the orthogonal decomposition of A is done

Orthogonal decomposition of A :

$$A = P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

need to determine $\perp P$, $\perp Q$ so :

$$P^T A Q = \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix}$$

We will illustrate algorithm on:

$$A = \begin{bmatrix} * & * & 0 \\ * & * & * \\ * & * & * \\ \square & * & * \end{bmatrix} \in \mathbb{R}^{4\times 3}$$

where : $*$ denotes arbitrary value
$\square$ -"- non-zero value
0 -"- zero value

Decomposition alg. starts by finding an $\perp P$ so :

$$\underbrace{P^T}_{\in \mathbb{R}^{4\times 4}} \underbrace{A}_{\in \mathbb{R}^{4\times 3}} = \text{an echelon matrix}$$

Echelon matrices

Staircase for a matrix

start of stair-case

$$M_1 = \begin{pmatrix} 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & \square & 0 & * & * \\ 0 & 0 & 0 & \square & * \end{pmatrix}$$

Rules :
staircase never goes up
goes down just enough so all entries below it, if any, are zero

Echelon matrix : has a staircase with
all steps (if any) = 1 row

e.g.

$M_1 \neq$ echelon

$M_2 =$ echelon

$$M_2 = \begin{pmatrix} \square & * & * & * & * \\ 0 & 0 & \square & * & * \\ 0 & 0 & 0 & 0 & \square \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\perp$ Transformation to echelon form

based on introducing zeros in columns

Putting a zero in a column

Can choose $\perp$ H, so

$$H\begin{bmatrix} a \\ b \\ c \\ \delta \\ \varepsilon \\ f \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ \square \\ 0 \\ f \end{bmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{- unchanged} \\ \leftarrow \text{non-zero} \\ \\ \leftarrow \text{unchanged} \end{array}$$

$0 \neq \varepsilon$

a, b, c, f ≡ unchanged but ε changed to zero

H has structure

$$H = \left[\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \multicolumn{2}{c|}{\bar{H}_{2\times 2}} & 0 \\ 0 & 0 & 0 & & & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$\bar{H}$ is the part of H which does the job of transforming $\binom{\delta}{\varepsilon}$ into $\binom{\square}{0}$. The identity matrices on the block diagonal of H are there so $(a\,b\,c)^T$ and f are unchanged by H.

Take:

$$\bar{H}_{2\times 2} = I_{2\times 2} - \frac{2uu^T}{\|u\|^2}$$

this is a so-called Givens rotation matrix

with

$$u \triangleq \binom{\delta}{\varepsilon} + e^1 \left\| \binom{\delta}{\varepsilon} \right\| \operatorname{sign}(\delta)$$

$$e^1 \triangleq \binom{1}{0} \quad ; \quad \operatorname{sign}(\delta) = \begin{cases} 1 & \text{if } \delta \geq 0 \\ -1 & \text{if } \delta < 0 \end{cases}$$

Does it zero the ε-element?

Check by direct calculation:

notation:

$$a \triangleq \binom{\delta}{\varepsilon}$$

$$\tilde{\delta} \triangleq \delta + \|a\| \operatorname{sign}(\delta)$$

$$uu^T = \binom{\tilde{\delta}}{\varepsilon}(\tilde{\delta}\ \varepsilon)$$

$$= \begin{bmatrix} \tilde{\delta}^2 & \varepsilon\tilde{\delta} \\ \varepsilon\tilde{\delta} & \varepsilon^2 \end{bmatrix}$$

$$\|u\|^2 = \delta^2 + \varepsilon^2 + 2|\delta| \left\| \binom{\delta}{\varepsilon} \right\| + \left\| \binom{\delta}{\varepsilon} \right\|^2$$

$$= 2(\|a\|^2 + |\delta|\,\|a\|)$$

$$\bar{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \tilde{\delta}^2 & \varepsilon\tilde{\delta} \\ \varepsilon\tilde{\delta} & \varepsilon^2 \end{bmatrix} \cdot \frac{1}{\|a\|^2 + |\delta|\,\|a\|}$$

$$\Rightarrow \quad \bar{H}\begin{pmatrix}\delta\\ \varepsilon\end{pmatrix} \triangleq \begin{pmatrix}v_1\\ v_2\end{pmatrix} = \begin{pmatrix}\bar{h}_{*1}\\ \underbrace{\bar{h}_{*2}}_{\boxed{\text{second row of } \bar{H}}}\end{pmatrix}\begin{pmatrix}\delta\\ \varepsilon\end{pmatrix}$$

$$\Rightarrow \quad v_2 = \bar{h}_{*2}\begin{pmatrix}\delta\\ \varepsilon\end{pmatrix}$$

$$= \frac{1}{\|a\|^2+|\delta|\|a\|}\left(-\varepsilon\tilde{\delta},\ \|a\|^2+|\delta|\|a\|-\varepsilon^2\right)\begin{pmatrix}\delta\\ \varepsilon\end{pmatrix}$$

$$= \frac{-\varepsilon\delta(\delta+\|a\|\operatorname{sign}(\delta))+\varepsilon\|a\|^2+\varepsilon|\delta|\|a\|-\varepsilon^3}{\|a\|^2+|\delta|\|a\|}$$

$$= \frac{-\varepsilon\delta^2-\varepsilon|\delta|\|a\|+\varepsilon\delta^2+\varepsilon^3+\varepsilon|\delta|\|a\|-\varepsilon^3}{\|a\|^2+|\delta|\|a\|}$$

$$\equiv 0 \ ! \qquad \boxed{\text{YES}}$$

Is $\bar{H}$ an orthogonal matrix ?

$$\bar{H}^T = I^T - \frac{2}{\|u\|^2}(uu^T)^T = I - \frac{2}{\|u\|^2}uu^T = \bar{H}$$

$$\Downarrow$$

$$\bar{H}^T\bar{H} = \left(I - \frac{2}{\|u\|^2}uu^T\right)\left(I - \frac{2}{\|u\|^2}uu^T\right)$$

$$= I \underbrace{- \frac{2}{\|u\|^2}uu^T - \frac{2}{\|u\|^2}uu^T + \frac{4}{\|u\|^2}\cdot\frac{(uu^T)(uu^T)}{\|u\|^2}}_{=0}$$

$$= I \qquad \boxed{\text{YES}}$$

Algorithm for $\perp$ transformation to echelon form

e.g. $A = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ \square & * & * \end{bmatrix} \neq$ echelon matrix because of the 4-row step for column 1

fix column 1 :

choose $\perp$ H_1 so

$$H_1 \underbrace{\begin{pmatrix} * \\ * \\ * \\ \square \end{pmatrix}}_{\boxed{\text{column 1 of A}}} = \begin{pmatrix} * \\ * \\ \square \\ 0 \end{pmatrix}$$

$$\boxed{H_1 = \begin{bmatrix} I_{2\times 2} & 0 \\ 0 & \bar{H} \end{bmatrix}}$$

$\boxed{\text{for the appropriate } \bar{H}}$

Then

$$H_1 A = H_1 [a_{*1}\ a_{*2}\ a_{*3}]$$
$$= [H_1 a_{*1} \mid H_1 a_{*2} \mid H_1 a_{*3}]$$

---

by reducing the 4-row step for column 1 to a 1-row step, using a sequence of $\perp$ transformations H_i to zap to zero the bottom 3 entries in column 1

$$H_1 A = \begin{bmatrix} * & * & * \\ * & * & * \\ \square & * & * \\ 0 & * & * \end{bmatrix} \triangleq A_1 \neq \text{echelon matrix}$$

changed

may be changed

choose $\perp$ H_2 so

$$H_2 \underbrace{\begin{pmatrix} * \\ * \\ \square \\ 0 \end{pmatrix}} = \begin{pmatrix} * \\ \square \\ 0 \\ 0 \end{pmatrix}$$

$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \bar{H} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ for the appropriate $\bar{H}$

Then

$$H_2 A_1 = \begin{bmatrix} * & * & * \\ \square & * & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix} \triangleq A_2 \neq \text{echelon matrix}$$

choose $\perp$ H_3 so

$$H_3 \underbrace{\begin{pmatrix} * \\ \square \\ 0 \\ 0 \end{pmatrix}} = \begin{pmatrix} \square \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad H_3 = \begin{bmatrix} \bar{H} & 0 \\ 0 & I_{2\times 2} \end{bmatrix}$$

$\}$ indicates those entries which are changed by the orthogonal transformation concerned.

So that:

$$H_3 A_2 = \begin{bmatrix} \square & * & * \\ 0 & * & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix} \triangleq A_3$$

Suppose actually that:

$$A_3 = \begin{bmatrix} \square & * & * \\ 0 & * & * \\ 0 & \square & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \text{echelon matrix}$$

because of 2-row step for Col. 2

zero by chance

fix column 2

choose $\perp$ H_4 so

$$H_4 \underbrace{\begin{pmatrix} * \\ * \\ \square \\ 0 \end{pmatrix}} = \begin{pmatrix} * \\ \square \\ 0 \\ 0 \end{pmatrix}$$

Then

$$H_4 A_3 = \begin{bmatrix} \square & * & * \\ 0 & \square & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{bmatrix} \triangleq A_4$$

still OK, as zeros induced by by earlier H_i not destroyed by later H_i.

Suppose, actually that :

$$A_4 = \begin{bmatrix} \square & * & * \\ 0 & \square & \square \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{echelon matrix}$$

$\boxed{\text{if not use extra } H_i \text{ to zap the offending non-zero entries}}$

Then

$$\underbrace{A_4}_{=\underbrace{H_4 A_3}_{=\underbrace{H_3 A_2}_{=\underbrace{H_2 A_1}_{=H_1 A}}}} = \text{echelon matrix}$$

i.e. $\underbrace{H_4 H_3 H_2 H_1}$ A = echelon matrix

$\boxed{\text{would be OK if this matrix was orthogonal is it ?}}$

$$(H_4 H_3 H_2 H_1)^T (H_4 H_3 H_2 H_1)$$

$$= H_1^T H_2^T H_3^T H_4^T H_4 H_3 H_2 H_1 = I$$

$\boxed{\text{because } H_i,\ i=1,..,4 \text{ are all } \perp}$

$\Rightarrow$ $\boxed{\text{YES},\ H_4 H_3 H_2 H_2 \triangleq P \text{ is } \perp}$

end of alg. for $\perp$ transf. to echelon form

$\perp$ Decomposition of A

e.g. $A = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ \square & * & * \end{bmatrix}$

① Apply to A the Alg. for $\perp$ transf. to echelon form.

Finds $\perp$ P so $P^T A = \text{e.g.} \begin{bmatrix} \square & * & * \\ 0 & \square & \square \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \triangleq \bar{A}$

② Consider

$$\bar{A}^T = \begin{bmatrix} \square & 0 & 0 & 0 \\ * & \square & 0 & 0 \\ * & \square & 0 & 0 \end{bmatrix}$$

③ Apply to $\bar{A}^T$: Alg. for $\perp$ transf. to echelon form

Finds $\perp Q$ so

$$Q^T \bar{A}^T = \begin{bmatrix} \boxed{1} & * & 0 & 0 \\ 0 & \boxed{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

this occurs because the $\perp$ transformation alg. first zeroes the bottom 2 entries in column 1 of $\bar{A}^T$, giving rise to $\boxed{1}$ in $Q^T\bar{A}^T$. Then it zeroes the resulting bottom entry of column 2, giving rise to $\boxed{2}$. The zeros on the right of $\bar{A}^T$ are not destroyed by those transformations, and so appear in $Q^T\bar{A}^T$

④ Transposing

$$\underbrace{\bar{A}Q}_{P^TA} = \begin{bmatrix} \boxed{1} & 0 & 0 \\ * & \boxed{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So

$$P^TAQ = \left[\begin{array}{cc:c} \boxed{1} & 0 & 0 \\ * & \boxed{2} & 0 \\ \hdashline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{c:c} \hat{A} & 0 \\ \hdashline 0 & 0 \end{array}\right]$$

$\hat{A}$ = invertible as : $\det[\hat{A}] = \boxed{1}\cdot\boxed{2} \neq 0$

Hence : have found $\perp P$, $\perp Q$ so

$$A = P \left[\begin{array}{c:c} \hat{A} & 0 \\ \hdashline 0 & 0 \end{array}\right] Q^T \qquad \text{as required.}$$

$\boxed{\text{invertible}}$

The method presented here works for any non-zero A, but two or more of the zero sub-matrices in $\begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix}$ may be absent, depending on the particular A considered.

<u>More algebra relevant to optimization</u>

Consider decomposed $A \in \mathbb{R}^{m \times n}$, as:

(9.1) $$A = P \begin{bmatrix} \hat{A}_{r \times r} & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

invertible

$\perp$ $\perp$

maybe with some 0-blocks missing

- useful for finding A^+
- useful for finding $\mathcal{R}[A]$, $\mathcal{N}[A] \triangleq \{x : Ax = 0\}$

useful later for constrained optimization

Aside: the so-called singular value decomposition yields (9.1) with diagonal $\hat{A}$

(9.2) <u>Th</u> For $A \in \mathbb{R}^{m \times n}$ of (9.1):

(i) $\mathcal{R}[A] = \mathcal{L}[\underbrace{p_{*1}, \dots, p_{*r}}_{\text{colums of } P}]$

form an orthonormal basis for $\mathcal{R}[A]$

(ii) $\mathcal{N}[A] = \mathcal{L}[\underbrace{q_{*(r+1)}, \dots, q_{*n}}_{\text{columns of } Q}]$

$= \mathcal{R}[H]$ where $H = [q_{*(r+1)}, \dots, q_{*n}]$

(9.3) all x's in $\mathcal{N}[A]$ are generated by $H\theta$ as θ varies over $\mathbb{R}^{n-r}$

If some of the 0-blocks in (9.1) are missing, the effects on Th. (9.2) are easy to figure out.

Proof of (i) uses:

(9.4) $\underline{\underline{\text{Lemma}}}$ If $B \in \mathbb{R}^{p \times p}$ is non-singular then $\mathcal{R}[B] = \mathbb{R}^p$

$\boxed{\text{Aside:}}$

> To show that:
> set X = set Y
> we have to show that
> $X \subset Y$ and $Y \subset X$
>
> To show that $X \subset Y$:
> show $(x \in X) \Rightarrow (x \in Y)$

<u>Proof of Lemma (9.4):</u>

Proof that $\mathcal{R}[B] \subset \mathbb{R}^p$:

$$\mathcal{R}[B] = \{ \underbrace{Bx}_{\in \mathbb{R}^p} : x \in \mathbb{R}^p \} \subset \mathbb{R}^p$$

Proof that $\mathbb{R}^p \subset \mathcal{R}[B]$:

$$y \in \mathbb{R}^p \Rightarrow y = B\underbrace{B^{-1}y}_{x} = Bx \quad \text{for some } x$$

$$\Rightarrow y \in \{ Bx : x \in \mathbb{R}^p \} = \mathcal{R}[B]$$

QED

Proof of (i):

$$\mathcal{R}[A] = \{ Ax : x \in \mathbb{R}^n \}$$

$$= \left\{ P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{Q^T x}_{y} : x \in \mathbb{R}^n \right\}$$

$\boxed{\text{$y$ varies over $\mathcal{R}[Q^T] = \mathbb{R}^n$ because Q^T = invertible}}$

$$= \left\{ P \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{y}_{= \binom{\tilde{y}}{\bar{y}}} : y \in \mathbb{R}^n \right\}$$

$$= \left\{ P \begin{bmatrix} \underbrace{\hat{A}\tilde{y}}_{z} \\ 0 \end{bmatrix} : \begin{pmatrix} \tilde{y} \\ \bar{y} \end{pmatrix} \in \mathbb{R}^n \right\}$$

$\boxed{\text{$z$ varies over $\mathcal{R}[\hat{A}] = \mathbb{R}^r$ - because $\hat{A}$ = invertible}}$

$$= \left\{ P \begin{pmatrix} z \\ 0 \end{pmatrix} : z \in \mathbb{R}^r \right\} = \left\{ \sum_{i=1}^{r} P_{*i} z_i : z \in \mathbb{R}^r \right\}$$

$$= \mathcal{L}[P_{*1}, \dots, P_{*r}]$$

QED

Orthogonal complement of a linear subspace

For a l.s.s. L of $\mathbb{R}^n$:

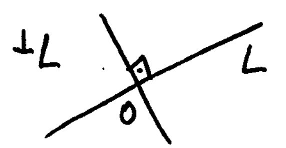
linear sub-space

$\boxed{\text{the orthogonal complement of } L}$

$^\perp L \triangleq$ the set of vectors which are orthogonal to every vector in L

$= \{ z \in \mathbb{R}^n : z^T x = 0, \ \forall x \in L \}$

in $\mathbb{R}^2$:



in $\mathbb{R}^3$:

(9.5) Th For a l.s.s. L of $\mathbb{R}^n$:

(i) $^\perp L$ is a l.s.s. of $\mathbb{R}^n$

(ii) $^\perp(^\perp L) = L$

Orthogonal projection

(9.6) Th Suppose L = a l.s.s. of $\mathbb{R}^n$

Then: every $z \in \mathbb{R}^n$ can be written as :

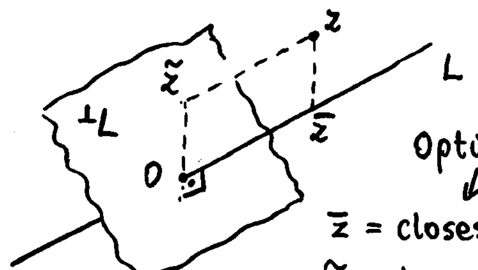
$$z = \underset{L}{\underset{\cap}{\bar{z}}} + \underset{^\perp L}{\underset{\cap}{\tilde{z}}}$$

the orthogonal projection of z on L

the orthogonal projection of z on $^\perp L$

for unique $\bar{z}, \tilde{z}$

in $\mathbb{R}^3$:



Optimization again!

$\bar{z}$ = closest pt. in L to z

$\tilde{z}$ = closest pt. in $^\perp L$ to z

(9.7) Th If $L = \mathcal{R}[A]$ for some $A \in \mathbb{R}^{m \times n}$

then:

$\bar{z} = AA^+ z$

$\tilde{z} = (I - AA^+) z$

(£)

A useful fact about $A \in \mathbb{R}^{m \times n}$

(9.8) **Th** $\quad {}^{\perp}\mathcal{R}[A] = \mathcal{N}[A^T]$

Proof that ${}^{\perp}\mathcal{R}[A] \subset \mathcal{N}[A^T]$:

$$z \in {}^{\perp}\mathcal{R}[A]$$

$\Rightarrow z^T y = 0 \,, \ \forall y \in \mathcal{R}[A]$

$\Rightarrow z^T A x = 0 \,, \ \forall x \in \mathbb{R}^n$

$\Rightarrow (A^T z)^T x = 0 \,, \ \forall x \in \mathbb{R}^n$

$\boxed{\text{choose } x = A^T z}$

$\Rightarrow (A^T z)^T A^T z = 0 \ \Rightarrow \|A^T z\|^2 = 0$

$\Rightarrow A^T z = 0$

$\Rightarrow z \in \mathcal{N}[A^T]$

Hence $\ {}^{\perp}\mathcal{R}[A] \subset \mathcal{N}[A^T]$

(f) closest point in $\mathcal{R}[A] = L$ to z:

$$\underbrace{\hat{x}}_{A^+ z} = \arg \min_{x \in \mathbb{R}^n} \|Ax - z\| \ \Rightarrow \ \bar{z} = A\hat{x} = AA^+ z$$

Proof that $\mathcal{N}[A^T] \subset {}^{\perp}\mathcal{R}[A]$:

$$z \in \mathcal{N}[A^T]$$

$\Rightarrow A^T z = 0$

$\Rightarrow x^T \underbrace{(A^T z)}_{y} = 0 \,, \ \forall x \in \mathbb{R}^n$

$\boxed{x^T y = \sum x_i y_i = \sum y_i x_i = y^T x}$

$\Rightarrow (A^T z)^T x = 0 \,, \ \forall x \in \mathbb{R}^n$

$\Rightarrow z^T \underbrace{A x}_{y} = 0 \,, \ \forall x \in \mathbb{R}^n$

varies all over $\mathcal{R}[A]$
as x varies over $\mathbb{R}^n$

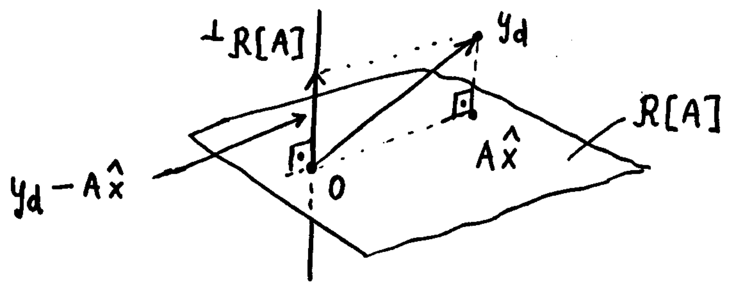
$\Rightarrow z^T y = 0 \,, \ \forall y \in \mathcal{R}[A]$

$\Rightarrow z \in {}^{\perp}\mathcal{R}[A]$

Hence $\mathcal{N}[A^T] \subset {}^{\perp}\mathcal{R}[A]$

QED

Application of Th. (9.8) to give increased insight for $\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2$



It is obvious geometrically that:

$y_d - A\hat{x} \perp$ to every vector in $\mathcal{R}[A]$

$\Rightarrow \; y_d - A\hat{x} \in {}^{\perp}\mathcal{R}[A] \underset{\boxed{\text{Th (9.8)}}}{=} \mathcal{N}[A^T]$

$\Rightarrow \; A^T(y_d - A\hat{x}) = 0$

$\Rightarrow \; A^T y_d = A^T A \hat{x}$

$\boxed{\text{assuming that } A^T A = \text{invertible}}$

$\Rightarrow \; \underset{\boxed{\text{unique}}}{\hat{x}} = (A^T A)^{-1} A^T y_d$

$\boxed{\text{computable formula for } \hat{x} \text{ !}}$

So: if $A^T A$ = invertible

then the unique global minimizer is:

$$\hat{x} = (A^T A)^{-1} A^T y_d$$

Since there is only one global minimizer, it must be the global minimizer of least norm.

Hence:

$$A^+ y_d = (A^T A)^{-1} A^T y_d \;, \; \forall y_d$$

$$\Rightarrow \; A^+ \underset{\boxed{\text{only if } (A^T A)^{-1} \text{ exists}}}{=} (A^T A)^{-1} A^T$$

- the above is a new derivation of a formula for A^+ for a special case, yielding new insight.

BUT: forming $A^T A$ undesirable numerically

$\Rightarrow$ formula of § 7 for A^+ is better as it works even if $A^T A \neq$ invertible and avoids calculation of $A^T A$.

Fixed point algorithms for solving $h(x)=0$

where : $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$

e.g. for solving optimality condition $\nabla V(x)=0$

potentially relevant alg. when
$V \neq$ quadratic and/or n = very big.

Idea : Rewrite $h(x)=0$ as $x = f(x)$

if possible

Then $(\hat{x} = \text{a solution}) \Leftrightarrow [h(\hat{x})=0]$

$$\Leftrightarrow [\hat{x} = f(\hat{x})]$$

so if $\hat{x}$ goes into f, then $\hat{x}$ comes out, i.e. $\hat{x}$ is a fixed-point of f

Suggests : "fixed-point" alg. :

$$x_{j+1} = f(x_j) \quad ; \quad x_0 \in \mathbb{R}^n \ , \ \forall j$$

Convergence of x_j to $\hat{x}$?
Stopping condition ?

Analysis needs :

Definition : f is called γ - Lipschitz (on $\mathbb{R}^n$) if :

$$\| f(x) - f(y) \| \leq \gamma \| x-y \| \ , \ \forall x,y \in \mathbb{R}^n$$

$\gamma \geq 0$

smallest possible γ will be best for us

If $\gamma < 1 \Rightarrow f$ is called a contraction

Eg. if $f(x) = a + Bx$, $a \in \mathbb{R}^n$, $B \in \mathbb{R}^{n \times n}$

$$\Rightarrow \| f(x) - f(y) \| = \| (a+Bx) - (a+By) \|$$
$$\leq \| B(x-y) \|$$
$$\leq \| B \| \ \| x-y \| \ , \ \forall x,y \in \mathbb{R}^n$$

$\Rightarrow$ can use $\gamma = \| B \|$, where :

$\| B \| = \| B \|_2$ = induced norm $\triangleq \sqrt{\lambda_{max}(B^T B)}$,
or
$\| B \| = \| B \|_F$ = Frobenius n. $\triangleq \sqrt{\sum_i \sum_j b_{ij}^2}$,

$\| B \|_F \geq \| B \|_2$ (so $\| B \|_F$ not best for γ)
BUT $\| B \|_F$ is much easier to compute

(10.1) $\underline{\underline{\text{Th}}}$ If f is a contraction then:

$$x = f(x)$$

<u>has a unique solution</u> $\hat{x}$.

<u>Proof : (only uniqueness</u>, existence harder)

non-unique solution

$\Downarrow$

$\exists$ solutions $\bar{x}$ and $\tilde{x}$ with $\bar{x} \neq \tilde{x}$

$\Downarrow$

$$\|\bar{x}-\tilde{x}\| = \|f(\bar{x})-f(\tilde{x})\|$$
$$\leq \underbrace{\gamma}_{\boxed{<1}} \|\bar{x}-\tilde{x}\|$$
$$< \|\bar{x}-\tilde{x}\| \leftarrow \boxed{\text{contradiction}}$$

Hence : unique solution.

QED

(10.2) $\underline{\underline{\text{Th}}}$ Consider : $x_{j+1} = f(x_j)$; $x_0 \in \mathbb{R}^n$, $\forall j$.

If f is γ-Lipschitz with $\gamma < 1$

(i.e. a contraction)

then:

$\exists$ a unique solution $\hat{x}$ of $x = f(x)$, and:

(i) $0 \leq \|x_j - \hat{x}\| \leq \gamma^j \|x_0 - \hat{x}\|$, $\forall j \geq 0$

so, since $\gamma < 1$

(ii) $x_j \to \hat{x}$ $\leftarrow$ $\boxed{\text{from (i): smaller } \gamma \text{ implies faster, guaranteed convergence}}$

<u>Proof of (i):</u>

$$\|x_{j+1} - \hat{x}\| = \|f(x_j) - f(\hat{x})\|$$
$$\leq \gamma \|x_j - \hat{x}\| \qquad (10.3)$$

so, iterating this,

$$\|x_j - \hat{x}\| \leq \gamma^j \|x_0 - \hat{x}\| , \forall j \geq 0$$

$\Rightarrow$ (i)

QED

A stopping condition

Suppose we would be satisfied with an x_j near enough to $\hat{x}$ in that:

$$\| x_j - \hat{x} \| < \varepsilon \leftarrow \boxed{\text{pre-specified e.g. } 10^{-8}}$$

Not directly useful as stopping-condition since $\hat{x}$ = unknown.

However:

(10.4) **Th** If f is γ-Lipschitz with $\gamma < 1$, then:

$$\| x_j - \hat{x} \| < \frac{\gamma}{1-\gamma} \| x_j - x_{j-1} \| \quad , \forall j > 1$$

So we can stop iterating, knowing that

$$\| x_j - \hat{x} \| < \varepsilon$$

if we continue iterating until

$$\frac{\gamma}{1-\gamma} \| x_j - x_{j-1} \| < \varepsilon$$

easy to test

Proof of Th (10.4):

i.e. of $\| x_j - \hat{x} \| \leq \frac{\gamma}{1-\gamma} \| x_j - x_{j-1} \|$ if $\gamma < 1$

Now:

$$\| x_{j-1} - \hat{x} \| = \| x_{j-1} - x_j + x_j - \hat{x} \|$$

$$\leq \| x_{j-1} - x_j \| + \underbrace{\| x_j - \hat{x} \|}_{\boxed{(10.3) \Rightarrow \ \leq \gamma \| x_{j-1} - \hat{x} \|}}$$

$$\leq \| x_{j-1} - x_j \| + \gamma \| x_{j-1} - \hat{x} \|$$

So:

$$\underbrace{(1-\gamma)}_{\boxed{>0 \text{ since } \gamma < 1}} \| x_{j-1} - \hat{x} \| \leq \| x_{j-1} - x_j \|$$

so

$$\| x_{j-1} - \hat{x} \| \leq \frac{1}{1-\gamma} \| x_{j-1} - x_j \|$$

Finally

$$\| x_j - \hat{x} \| \underset{\boxed{(10.3)}}{\leq} \gamma \| x_{j-1} - \hat{x} \|$$

$$\leq \frac{\gamma}{1-\gamma} \| x_{j-1} - x_j \|$$

$$= \frac{\gamma}{1-\gamma} \| x_j - x_{j-1} \| \qquad \text{QED}$$

The fixed-point alg. can now be coded as:

- x_{new} := initial estimate x_0 of $\hat{x}$

Repeat:

- $x_{old} := x_{new}$
- $x_{new} := f(x_{new})$

Until : $\frac{\gamma}{1-\gamma} \| x_{new} - x_{old} \| < \varepsilon$

Question: does it ever terminate?
Answer: yes – always if only $\gamma < 1$.

About rewriting $h(x)=0$ as $x = f(x)$ for the linear case

$$0 = h(x) \triangleq Ax - b \quad , A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n$$

Methodology:

split A as $A = K + E$

easily invertible approximation to A

error

Then: $0 = Ax - b$
$= Kx + Ex - b$
$\Rightarrow \quad Kx = b - Ex$
$\Rightarrow \quad x = K^{-1}[b - Ex]$

the re-written $h(x)=0$

So:

$$\left\{ \begin{matrix} \text{the split} \\ A = K+E \end{matrix} \right\} \Rightarrow x = K^{-1}[b - Ex] \triangleq f(x)$$

$$\Downarrow$$

$$f = \underbrace{\| K^{-1} E \|}_{\gamma} - \text{Lipschitz}$$

Different choices for K

$\Downarrow$

different $f(x)$'s and different fixed-point algs.

$x_{j+1} = f(x_j)$

$\Downarrow$

$\| x_j - \hat{x} \| \leq \gamma^j \| x_0 - \hat{x} \|$ for different γ's

$\Downarrow$

try to choose easily invertible K, so

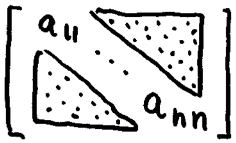
γ = small (est)

$\boxed{\text{best choice depends on } A}$

<u>Some famous choices for K</u>

<u>for the important case when : $a_{ii} \neq 0$</u>, $\forall i$

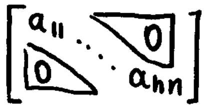
View A as:

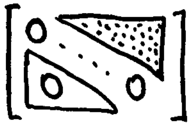
$A =$ 

$\equiv$

$L + D + U \longrightarrow$

$L =$ 

$D =$ 

$U =$ 

We want: $A = K + E$

$\boxed{\text{easily invertible approx. to } A}$

<u>Jacobi iteration</u>:

choose $K = D$, $E = L + U$

$\boxed{\text{diagonal = easily invertible}}$

<u>Gauss-Seidel iteration:</u>

choose: $K = L + D$, $E = U$

$\boxed{\text{lower-triangular = easily invertible}}$

<u>Relaxation iteration:</u>

choose: $K = K_\omega \triangleq L + \frac{D}{\omega}$ $\boxed{\text{scalar}}$

$\Rightarrow E = E_\omega \triangleq -\frac{D}{\omega} + D + U$

giving: $\gamma_\omega = \| K_\omega^{-1} E_\omega \|$

and try to choose ω so γ_ω = small

ω = so-called relaxation factor

$\omega > 1$: over-relaxation

$\omega < 1$: under-relaxation

$\omega = 1$: relaxation = Gauss-Seidel

Final remarks on fixed-point algs. :

- For a given A , sometimes it might not be possible to find a K so that

  $\| K^{-1} E \| = \gamma$ is less than one.

  However, a suitable K exists sufficiently often for fixed-point algorithms to be of interest.

- fixed-point algs. converge to $\hat{x}$ if only $f = \gamma$-Lipschitz and $\gamma < 1$

  may be linear or nonlinear

- fixed-pt. algs. = simple algs.
- require little storage

$\Big\}$ GOOD

BUT

- fixed-pt. algs. fail (almost always) if $\gamma > 1$

Now :
- a little more of unconstrained opt.
- start of equality-constrained opt.

More notation:

Consider $h : \mathbb{R}^n \to \mathbb{R}^m$ $\leftarrow$ $\boxed{h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}}$

$$h(x+\delta x) \cong h(x) + \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial h_1}{\partial x_i} \delta x_i \\ \cdots \\ \sum_{i=1}^{n} \frac{\partial h_m}{\partial x_i} \delta x_i \end{bmatrix}$$

$$\equiv h(x) + \underbrace{\begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \cdots & & \cdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}}_{\triangleq h_x(x) \text{ - jacobian of } h} \begin{bmatrix} \delta x_1 \\ \vdots \\ \delta x_n \end{bmatrix}$$

i.e.

$$h(x+\delta x) \cong h(x) + h_x(x)\delta x$$

validity requires all $\frac{\partial h_j}{\partial x_i}$ to exist and be continuous near x; guaranteed if h is a C^1 function meaning all 1st-order p.d.s. exist and are continuous on $\mathbb{R}^n$.

Optimization for solving : $h(x) = 0$ $\boxed{\text{maybe nonlinear}}$

Could solve using :

- a nonlinear vector-equation solver alg.

or by

- trying to choose x so $\|h(x)\|^2 = 0$

  i.e. solving : $\min_{x \in \mathbb{R}^n} \|h(x)\|^2$ (#)

  i.e. $\min_{x \in \mathbb{R}^n} V(x)$ with $V(x) \triangleq \|h(x)\|^2$

Formula for $\nabla V(x)$ etc.

$$V(x) = \|h(x)\|^2 = \sum_{i=1}^{n} h_i(x)^2$$

so

$$\frac{\partial V}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial}{\partial x_j} h_i^2(x) = \sum_{i=1}^{n} 2 \frac{\partial h_i}{\partial x_j}(x)$$

(#) this approach useful e.g. when fixed-point algs. cannot be used because we cannot get $\gamma < 1$

So:

$$\nabla V(x) = \begin{pmatrix} \frac{\partial V}{\partial x_1} \\ \vdots \\ \frac{\partial V}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} 2 \frac{\partial h_i}{\partial x_1} h_i \\ \vdots \\ \sum_{i=1}^{n} 2 \frac{\partial h_i}{\partial x_n} h_i \end{pmatrix}(x)$$

$$= 2 \underbrace{\begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_n}{\partial x_n} \end{bmatrix}}_{h_x(x)^T} \overbrace{\begin{pmatrix} h_1(x) \\ \vdots \\ h_n(x) \end{pmatrix}}^{h(x)}$$

So:

(11.1) $\nabla V(x) = 2\, h_x(x)^T h(x)$

Similarly, it can be calculated that

(11.2) $V_{xx}(x) = 2\, h_x(x)^T h_x(x) + 2 \sum_{i=1}^{n} h_i(x)\, h_{i\,xx}(x)$

$h_i \to 0$ when $V = \sum h_i^2$ decreases to zero

evaluation = expensive

So:

(11.2) $V_{xx}(x) \cong 2\, h_x(x)^T h_x(x)$

with the approximation getting better as $x \to$ a solution of $h(x) = 0$ since then $h_i \to 0$.

Hence: we can use the approximation

$$\bar{P}(x) = 2\, h_x(x)^T h_x(x) \quad \text{to } V_{xx}(x)$$

in a Newton-type alg.

In fact: $\bar{P}(x) \geq 0$

but is not necessarily > 0

we can, however, approximate $\bar{P}(x)$ by a positive-definite $P(x)$ by changing all zero λ_i of $\bar{P}$ to $\varepsilon > 0$

$\Rightarrow$ we can get a positive-definite approximation $P(x)$ to $V_{xx}(x)$

Once we have an approximation $P(x) > 0$ to $V_{xx}(x)$, $\forall x$ we can apply the P-alg. of § 4 to minimize $V(x)$.

A iteration j of the P-alg.:

$$s_j = -P(x_j)^{-1} \nabla V(x_j)$$

= the (so-called) <u>Gauss-Newton search direction</u>

Using this s_j at each iteration yields the Gauss-Newton Alg. for solving $h(x) = 0$ by minimizing $\min_{x \in \mathbb{R}^n} \|h(x)\|^2$

Equality-constrained optimization

$\min_{x \in F} V(x)$ where $F = \{x \in \mathbb{R}^n : h(x) = 0\}$

vector equality constraint
$h : \mathbb{R}^n \to \mathbb{R}^m$
usually $m \leq n$ to help ensure that $F \neq \emptyset$

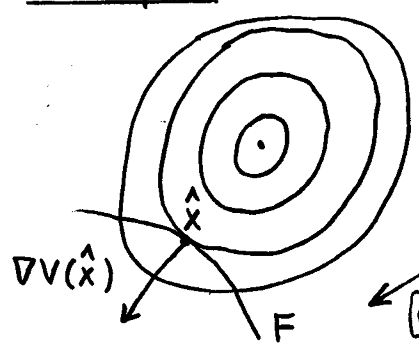
Equivalent problem formulations:

$$\min_{x \in \mathbb{R}^n} V(x) \quad \text{subject to } h(x) = 0$$

or

$$\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = 0\}$$

Example



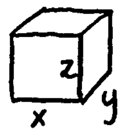
For unconstrained opt.:
$\nabla V(\hat{x}) = 0$

For constrained opt.:
$\nabla V(\hat{x}) \neq 0$

as shown here

Example

box



maximize volume

subject to : area of material = A (given to us)

i.e.

$$\max_{(x,y,z)\in\mathbb{R}^3} xyz$$

subject to $2(xy+yz+xz) = A$

i.e. since $\max V$ is eqivalent to $\min(-V)$

$$\Rightarrow \min_{(x,y,z)\in\mathbb{R}^3} \{-xyz : 2(xy+yz+xz) = A\}$$

Example: $x_{k+1} = A x_k + B u_k$, $x_0 = x_0$

Minimize control effort needed to reach a desired state x_d at a time N , i.e.

$$\min_{(u_0,\dots,u_{N-1})\in\mathbb{R}^N} (u_0^2 + u_1^2 + \dots + u_{N-1}^2)$$

subject to :

$$x_d = A^N x_0 + A^{N-1} B u_0 + \dots + B u_{N-1}$$

Linear equality-constrained opt. of general V :

$$\min_{x\in\mathbb{R}^n} \{ V(x) : Dx = f \}$$

D: a m×n matrix with linearly independent rows (to avoid redundant constraints)

$Dx = f$: general linear equality constraint

(I) Simple elimination method

e.g. $\min_{x\in\mathbb{R}^2} \{ V(x) : a x_1 + b x_2 = f \}$

$\Downarrow$ — solving for x_2

$$x = \begin{pmatrix} x_1 \\ \frac{f - a x_1}{b} \end{pmatrix} \in F , \forall x_1$$

$\Downarrow$

can solve this constrained problem by:

$\boxed{\text{unconstr. min.}} \rightarrow \min_{x_1\in\mathbb{R}} V\left(\begin{pmatrix} x_1 \\ (f-a x_1)/b \end{pmatrix}\right) \leftarrow$ EASY

this method = OK only for simple constraints.

$\Rightarrow$ a better approach is needed

(II) <u>The null-space method</u>

based on a deeper understanding of F
$$\{x \in \mathbb{R}^n : Dx = f\}$$

(11.3) <u>Th.</u> $F \neq \emptyset$ iff $DD^+f = f$

easy to test

If $F \neq \emptyset$ then there are feasible x's so we can proceed to $\min_{x \in F} V(x)$

If $F = \emptyset$ then there are no feasible x's so we cannot $\min_{x \in F} V(x)$

$\Rightarrow$ need to re-think the constraints

<u>Proof of (11.3):</u>

$F \neq \emptyset$ iff $\exists\, x$ such that $f = Dx$

i.e. iff $f \in \{Dx : x \in \mathbb{R}^n\} = \mathcal{R}[D]$

i.e. iff $f \in \mathcal{R}[D]$

$$\| \quad \text{orthogonal projection Th. (9.6)}$$

$$\underset{\mathcal{R}[D]}{\bar{f}} + \underset{{}^{\perp}\mathcal{R}[D]}{\tilde{f}}$$

i.e. iff $f = \bar{f} = DD^+f$

Th (9.7)

So $F \neq \emptyset$ iff $f \in \mathcal{R}[D]$,

i.e. iff $f = DD^+f$

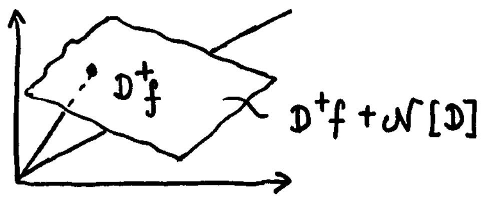
QED

More about F

(11.4) **Th** If $F \neq \emptyset$:

Then $F = D^+f + \mathcal{N}[D]$

$\stackrel{\Delta}{=} \{ D^+f + y : y \in \mathcal{N}[D] \}$



Proof: Assume $F \neq \emptyset$

$\xRightarrow{\boxed{\text{Th }(11.3)}} \; D\underbrace{D^+f}_{x} = f \;\Rightarrow\; D^+f \in F$ ← #

Proof that $D^+f + \mathcal{N}[D] \subset F$

$x \in D^+f + \mathcal{N}[D]$

$\Rightarrow x = D^+f + y$ for some $y \in \mathcal{N}[D]$

$\Rightarrow Dx = \underbrace{DD^+f}_{f\ (\#)} + \underbrace{Dy}_{0} = f$

Hence $x \in F$. QED

Proof that $F \subset D^+f + \mathcal{N}[D]$:

$x \in F \Rightarrow Dx = f \stackrel{\#}{=} DD^+f$

$\Rightarrow D(\underbrace{x - D^+f}_{y \in \mathcal{N}[D]}) = 0$

$\Rightarrow x - D^+f = y$ for some $y \in \mathcal{N}[D]$

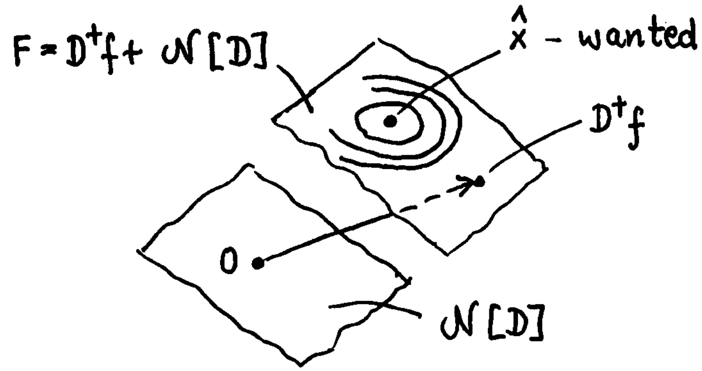
$\Rightarrow x = D^+f + y$ for some $y \in \mathcal{N}[D]$

$\Rightarrow x \in \{ D^+f + y : y \in \mathcal{N}[D] \}$

$\equiv D^+f + \mathcal{N}[D]$ QED

So, by Th. (11.4), if $\underline{F \neq \emptyset}$ we have this situation:

Ⓐ



Now we need an H such that:

$\mathcal{N}[D] = \mathcal{R}[H] \leftarrow$ to make easy calculations with $\mathcal{N}[D]$

Ⓑ <u>Finding an H so that $\mathcal{N}[D] = \mathcal{R}[H]$</u>

- Apply orthogonal decomposition alg. to D:

– If:
$$\left.\begin{array}{l} D = P\hat{D}Q^T \\ \text{or} \\ D = P\begin{bmatrix}\hat{D}\\ 0\end{bmatrix}Q^T \end{array}\right\} \Rightarrow \mathcal{N}[D] = \{0\}$$

not very likely

– If:
$$\left.\begin{array}{l} D = P\begin{bmatrix}\hat{D}_{r\times r} & 0\\ 0 & 0\end{bmatrix}Q^T \\ \text{or} \\ D = P\begin{bmatrix}\hat{D}_{r\times r}\\ 0\end{bmatrix}Q^T \end{array}\right\}$$

more likely

then:
$$\mathcal{N}[D] = \mathcal{R}[\underbrace{H}_{[q_{*(r+1)} \cdots q_{*n}]}]$$

So:
$$F = D^+f + \underset{\mathcal{R}[H]}{\underset{|||}{\mathcal{N}[D]}}$$
$$= \{D^+f + y : y \in \mathcal{R}[H]\}$$
$$= \{D^+f + H\theta : \theta \in \mathbb{R}^{n-r}\}$$

↑ very nice formula for a feasible set F

$\Downarrow$

Hence:

$$\min_{x \in F} V(x) = \min_{x \in \{D^+ f + H\theta \,:\, \theta \in \mathbb{R}^{n-r}\}} V(x)$$

constrained optimization

$$= \min_{\theta \in \mathbb{R}^{n-r}} V(D^+ f + H\theta) \qquad (\pounds)$$

unconstrained optim.
use C.G., Secant, etc.

Ⓒ Solving ($\pounds$) ... EASY

and then : an $\hat{x} = D^+ f + H\hat{\theta}$
for any g.m. $\hat{\theta}$ of ($\pounds$).

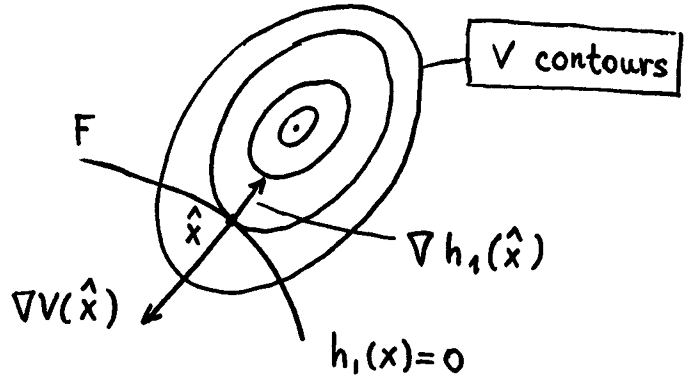
nice, easy method for solving linear equality-constrained opt. problems, by turning them into unconstrained opt. problems, by making use of the algebraic structure of F, namely using : $F = D^+ f + \mathcal{R}[H]$

Nonlinear equality constrained optimization – necessary conditions for optimality

$$\min_{x \in \mathbb{R}^n} \{ V(x) : h(x) = 0 \} \qquad (E)$$

$h : \mathbb{R}^n \to \mathbb{R}^m$ — potentially nonlinear

e.g.
$m = 1$
$h = h_1$



$$\nabla V(\hat{x}) = -2\, \nabla h_1(\hat{x})$$

i.e. $\nabla V(\hat{x}) = -\lambda_1 \nabla h_1(\hat{x})$ for some $\lambda_1 \in \mathbb{R}$ (12.0)

necessary condition for optimality of $\hat{x}$ in (E) for the case when $m = 1$

(12.1) **Th** $\hat{x}$ is a local minimizer for (E)

Then
$$\begin{cases} \nabla V(\hat{x}) = -\sum_{i=1}^{m} \lambda_i \nabla h_i(\hat{x}) \\ \text{for some } \lambda_i \in \mathbb{R},\ i=1,\dots,m \\ \text{provided } \nabla h_1(\hat{x}), \dots \nabla h_m(\hat{x}) \\ \text{are linearly independent} \end{cases}$$

a constraint qualification, usually satisfied, so not unduly restrictive

a necessary condition for equality-constrained optimality (a general version of (12.0))

More notation

$$V_x(x) \triangleq \left[\frac{\partial V}{\partial x_1}, \dots, \frac{\partial V}{\partial x_n}\right] = \nabla V(x)^T$$

$$h_x(x) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix}$$

$$\Rightarrow h_x(x)^T = [\nabla h_1(x) \dots \nabla h_m(x)]$$

Proof: Since $\hat{x}$ = a l.m. $\Rightarrow$

$V(\hat{x}+\delta x) \geqslant V(\hat{x})$, $\forall$ small δx such that $h(\hat{x}+\delta x) = 0$

$$h(\hat{x}+\delta x) \cong \underbrace{h(\hat{x})}_{0} + h_x(\hat{x})\delta x \qquad (\#)$$

define: $\hat{h}_x \triangleq h_x(\hat{x})$

$\Rightarrow V(\hat{x}+\delta x) \geqslant V(\hat{x})$ $\forall$ small δx such that $\hat{h}_x \delta x = 0$

for subtle reasons, the accuracy of expansion (#) is not sufficient for this $\Rightarrow$ to be valid unless the $\nabla h_i(\hat{x})$ are linearly independent

but $V(\hat{x}+\delta x) \geqslant V(\hat{x})$ $\forall \begin{cases} \text{small } \delta x \text{ s.t} \\ \hat{h}_x \delta x = 0 \end{cases}$

$$V(\hat{x}+\delta x) \cong V(\hat{x}) + \nabla V(\hat{x})^T \delta x$$

for small δx

Hence :

$$\nabla V(\hat{x})^T \delta x \geqslant 0 \quad \forall \begin{cases} \text{small } \delta x \text{ such that} \\ \hat{h}_x \delta x = 0 \end{cases}$$

i.e., since all is now linear, then :

$$\nabla V(\hat{x})^T \delta x \geqslant 0 \quad \forall\ \delta x \text{ such that } \hat{h}_x \delta x = 0$$

$$\Rightarrow \nabla V(\hat{x})^T \delta x \geqslant 0 \quad \forall\ \delta x \in \mathcal{N}[\hat{h}_x]$$

$\boxed{\text{since } (\hat{h}_x \delta x = 0) \Rightarrow (\hat{h}_x(-\delta x) = 0)}$

$$\Longrightarrow \left.\begin{array}{l} \nabla V(\hat{x})^T \delta x \geqslant 0 \\ \nabla V(\hat{x})^T (-\delta x) \geqslant 0 \end{array}\right\} \forall\ \delta x \in \mathcal{N}[\hat{h}_x]$$

$$\Longrightarrow \left.\begin{array}{l} \nabla V(\hat{x})^T \delta x \geqslant 0 \\ \nabla V(\hat{x})^T \delta x \leqslant 0 \end{array}\right\} \forall\ \delta x \in \mathcal{N}[\hat{h}_x]$$

$$\Longrightarrow \nabla V(\hat{x})^T \delta x = 0 \quad , \forall\ \delta x \in \mathcal{N}[\hat{h}_x]$$

$$\Longrightarrow \nabla V(\hat{x}) \in {}^{\perp}\mathcal{N}[\hat{h}_x] = \mathcal{R}[\hat{h}_x^T]$$

$\boxed{\text{easy to prove}}$

Hence

$$\nabla V(\hat{x}) = \hat{h}_x^T \theta \quad \text{for some } \theta \in \mathbb{R}^m$$

$$= \hat{h}_x^T (-\lambda) \quad \text{for some } \lambda \in \mathbb{R}^m$$

$$= [\nabla h_1(\hat{x}) \ldots \nabla h_m(\hat{x})](-\lambda)$$

$$\text{for some } \lambda \in \mathbb{R}^m$$

$$= \sum_{i=1}^{m} \nabla h_i(\hat{x}) \lambda_i$$

QED

our necessary condition for equality constrained optimality at $\hat{x}$;
we also need $h(\hat{x}) = 0$, of course.

note that this condition is valid if we assume, additionally, that
h & V are C^1 functions
so the 1st-order p.ds. and expansions are OK.

The Lagrangian and its role in the necessary conditions for optimality

The Lagrangian, namely:

$$\mathcal{L}(x,\lambda) \triangleq V(x) + \underbrace{\sum_{i=1}^{m} \lambda_i}_{\boxed{\text{Lagrange multipliers}}} h_i(x) \quad , \forall x$$

$$= V(x) + \underbrace{\lambda^T}_{\boxed{\text{Lagrange multiplier vector}}} h(x) \quad , \forall x$$

some use "−" here instead of a "+", so $\lambda_{theirs} = -\lambda_{mine}$

plays an important role in constrained opt.

Aside: notation ($\mathcal{L}$ is a f^n of x & λ):

$$\nabla_x \mathcal{L} \triangleq \left(\frac{\partial \mathcal{L}}{\partial x_1}, \ldots, \frac{\partial \mathcal{L}}{\partial x_n} \right)^T \triangleq \mathcal{L}_x^T$$

$$\nabla_\lambda \mathcal{L} \triangleq \left(\frac{\partial \mathcal{L}}{\partial \lambda_1} \ldots \frac{\partial \mathcal{L}}{\partial \lambda_n} \right)^T = \mathcal{L}_\lambda^T$$

Using $\mathcal{L}$ Th. (12.1) can be rewritten as:

(13.0) **Th** $\hat{x}$ is a local minimizer for (E)

Then $\nabla_x \mathcal{L}(\hat{x},\lambda) = 0$ for some $\lambda \in \mathbb{R}^m$

i.e. $\underbrace{\mathcal{L}(\hat{x},\lambda)_x}_{\boxed{\text{neat notation}}} = 0$ for some $\lambda \in \mathbb{R}^m$

Proof: $\mathcal{L}(x,\lambda) = V(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$

$\Rightarrow \mathcal{L}(\hat{x},\lambda)_x = V(\hat{x})_x + \sum_{i=1}^{m} \lambda_i h_i(\hat{x})_x$

$\Rightarrow \mathcal{L}(\hat{x},\lambda)_x^T = V(\hat{x})_x^T + \sum_{i=1}^{m} \lambda_i h_i(\hat{x})_x^T$

$\Rightarrow \nabla_x \mathcal{L}(\hat{x},\lambda) = \nabla_x V(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(\hat{x})$

$\underset{\boxed{\text{Th. (12.1)}}}{=} 0$ for some $\lambda \in \mathbb{R}^m$ if $\hat{x}$ = a local min. for (E)

QED

(13.1) **Th** (A property of L)

$$V(x) = L(x,\lambda) \ , \ \forall \lambda \in \mathbb{R}^m$$

if x is feasible ($x \in F$)

Proof: $L(x,\lambda) = V(x) + \lambda^T \underbrace{h(x)}_{=0} = V(x)$, $\forall \lambda$

$\boxed{\text{if } x \text{ is feasible}}$

QED

This property useful in proving the following sufficient condition for optimality of $\hat{x}$ for (E):

(13.2) **Th** $\hat{x}$ is a local minimizer for (E) if:

(i) $h(\hat{x}) = 0$

(ii) $L(\hat{x},\hat{\lambda})_x = 0$ for some $\hat{\lambda} \in \mathbb{R}^m$

and either:

(iii) $L(\hat{x},\hat{\lambda})_{xx} > 0$

or

(iv) $H^T L(\hat{x},\hat{\lambda})_{xx} H > 0$ $\boxed{\text{less restrictive than (iii)}}$

where H has linearly indep. columns

and $\mathcal{R}[H] = \mathcal{N}[h(\hat{x})_x]$

Notes on Th. (13.2):

- linear independence of the $\nabla h_i(\hat{x})$ is not needed
- V, h_i, $\forall i$, are assumed to be C^2 functions

$\boxed{\text{meaning all } \frac{\partial^2}{\partial x_i \partial x_j} \text{ exist and are continuous on } \mathbb{R}^n \text{, so a 2-nd order expansion is valid in the proof}}$

- $L(\hat{x},\hat{\lambda})_x = V(\hat{x})_x + \sum_{i=1}^m \hat{\lambda}_i h_i(\hat{x})_x$

$$= \left[\nabla V(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \nabla h_i(\hat{x})\right]^T$$

- $L(\hat{x},\hat{\lambda})_{xx} = \left[V(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i h_i(\hat{x})\right]_{xx}$

$$= V_{xx}(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \underbrace{h_i(\hat{x})_{xx}}_{\equiv\ \boxed{h_{i_{xx}}(\hat{x})}}$$

How can one find H ?

- one way to find a suitable H is to apply to $h(\hat{x})_x$ our orthogonal decomposition alg. giving:

$$h(\hat{x})_x = \underbrace{P}_{\perp} \begin{bmatrix} \hat{A}_{r\times r} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{Q^T}_{\perp}$$

from which : $H = [\underbrace{q_{*(r+1)} \cdots q_{*n}}_{\boxed{\text{orthonormal and hence linearly independent}}}]$

and $\mathcal{R}[H] = \mathcal{N}[h(\hat{x})_x]$

Proof of Th (13.2) :

$\boxed{\text{case } L(\hat{x},\hat{\lambda})_{xx} > 0}$

For small δx such that $h(\hat{x}+\delta x) = 0$:

$$V(\hat{x}+\delta x) \underset{\boxed{\text{Th } 13.1}}{=} L(\hat{x}+\delta x, \hat{\lambda})$$

$$\triangleq \underbrace{L(\hat{x},\hat{\lambda})}_{= V(\hat{x})} + \underbrace{L(\hat{x},\hat{\lambda})_x}_{= 0 \;\; \boxed{(ii)}} \delta x + \frac{1}{2}\delta x^T L(\hat{x},\hat{\lambda})_{xx} \delta x$$

$\boxed{(i) \wedge \text{Th } (13.1)}$

$$= V(\hat{x}) + \underbrace{\frac{1}{2}\delta x^T \underbrace{L(\hat{x},\hat{\lambda})_{xx}}_{>0} \delta x}_{>0 \, , \; \forall \delta x \neq 0}$$

$$> V(\hat{x}) \quad , \; \forall \; \delta x \neq 0 \text{ s.t. } h(\hat{x}+\delta x) = 0$$

$\Rightarrow \hat{x}$ = a local minimizer for (E)

$$\boxed{\begin{array}{l}\text{Case } L(\hat{x},\hat{\lambda})_{xx} \not> 0 \text{ but with} \\ \quad H^T L(\hat{x},\hat{\lambda})_{xx} H > 0\end{array}}$$

Since $\hat{x}$ = feasible

$\Rightarrow$ $\forall$ sufficiently small $\delta x \neq 0$ such that $h(\hat{x}+\delta x) = 0$:

$$0 = h(\hat{x}+\delta x) \cong \underbrace{h(\hat{x})}_{=0} + h_x(\hat{x})\delta x$$

$\Rightarrow$ so, for all such $\delta x \neq 0$: $h_x(\hat{x})\delta x = 0$

$\Rightarrow$ $\delta x \in \mathcal{N}[h_x(\hat{x})] = \mathcal{R}[H]$

$\Rightarrow$ $\delta x = H\theta$ for some θ

$\theta \neq 0$ — $\boxed{\text{since } \delta x \neq 0}$

Therefore,

$$V(\hat{x}+\delta x) \cong V(\hat{x}) + \tfrac{1}{2}\delta x^T L(\hat{x},\hat{\lambda})_{xx}\,\delta x$$

$$= V(\hat{x}) + \underbrace{\tfrac{1}{2}\theta^T H^T L(\hat{x},\hat{\lambda})_{xx} H\theta}_{>0 \text{ since } \theta \neq 0}$$

so

$V(\hat{x}+\delta x) > V(\hat{x})$, $\forall$ small $\delta x \neq 0$ s.t. $h(\hat{x}+\delta x) = 0$

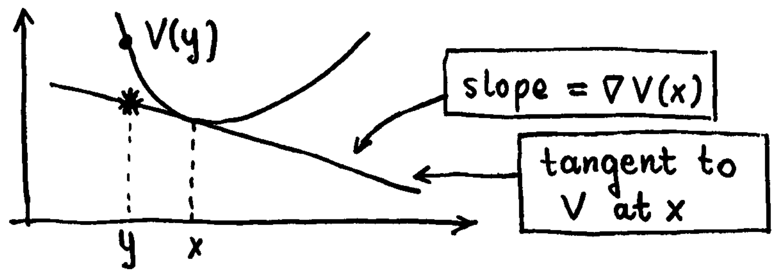
i.e. $\hat{x}$ = a l.m. of (E)

QED

More about convex functions

(13.3) **Th** For a convex C^1 function $V: \mathbb{R}^n \to \mathbb{R}$:

$$V(y) \geq \underbrace{V(x) + \nabla V(x)^T(y-x)}_{*} \quad , \forall\, x, y \in \mathbb{R}^n$$



i.e. $V(y)$ lies above its tangent at x, $\forall y$, $\forall x$

e.g. V = standard quadratic on $\mathbb{R}^n$

$$V(y) = V(x) + \nabla V(x)^T(y-x) + \underbrace{\tfrac{1}{2}(y-x)^T C(y-x)}_{\geq 0}$$

$$\geq V(x) + \nabla V(x)^T(y-x) \quad , \forall\, x, y \in \mathbb{R}^n$$

Special case of sufficient condition for optimality

(13.4) **Th** If $h(x) = Dx - f$

$\boxed{\text{linear equality constraints}}$

V = a convex C^1 function

Then : conditions (i), (ii), (iii) of Th. (13.2) guarantee that :

$\hat{x}$ = a global minimizer for (E)

$\boxed{\text{NOT just the local minimizer predicted by Th (13.2)}}$

Methodology for using Th. (13.2) to find a local minimizer for (E)

Th. (13.2) says (in simplest case):

$\hat{x}$ = a local minimizer for (E)

if $L(\hat{x}, \hat{\lambda})_x = 0$ for some $\hat{\lambda} \in \mathbb{R}^m$

$h(\hat{x}) = 0$; $L(\hat{x}, \hat{\lambda})_{xx} > 0$

Methodology is :

1. choose $x(\lambda)$ so $L(x(\lambda), \lambda)_x = 0$
2. find $\Lambda = \{\lambda : h(x(\lambda)) = 0\}$
3. find $\tilde{\Lambda} = \{\lambda \in \Lambda : L(x(\lambda), \lambda)_{xx} > 0\}$

   then $\tilde{X} = \{x(\lambda) : \lambda \in \tilde{\Lambda}\}$

   = a set of local minimizers,

   so choose and use an $x \in \tilde{X}$ giving least cost.

So, methodology for using Th. (13.2) delivers a set $\tilde{\Lambda}$ such that:

$$\forall \lambda \in \tilde{\Lambda}: \quad L[x(\lambda),\lambda]_x = 0 \ , \ h(x(\lambda)) = 0$$
$$L[x(\lambda),\lambda]_{xx} = 0$$

$\underset{\boxed{\text{Th (13.2)}}}{\Rightarrow}$ $x(\lambda)$ is a local minimizer for (E).

$\Rightarrow$ $\tilde{X}$ = set of local minimizers

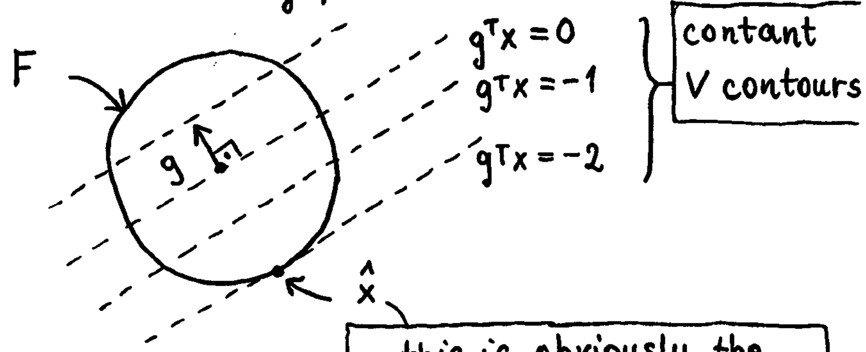
$\Rightarrow$ $x(\lambda) \in \tilde{X}$ yielding least $V \underset{\boxed{\text{hopefully}}}{=}$ global minim.

> in some cases it will not be possible to follow the above methodology exactly
> - but the approach should be applied as far as possible.
>
> Also:
> the methodology is usually practicable, for nonlinear h, only when $m = 1$, i.e. when there is ONE scalar equality constraint.

<u>Example</u>

$$\min_{x \in \mathbb{R}^2} \{ \underbrace{g^T x}_{V(x)} : \underbrace{x_1^2 + x_2^2 - 4}_{h_1(x)} = 0 \}$$

$g \neq 0$



> this is obviously the only local minimizer in this simple example
> - does our methodology find it?

The Lagrangian:

$$L[x,\lambda_1] = V(x) + \lambda_1 h_1(x)$$
$$= g^T x + \lambda_1 (x_1^2 + x_2^2 - 4)$$

① choose $x(\lambda_1)$ so $L[x(\lambda_1),\lambda_1]_x = 0$

$$L[x,\lambda]_x = \left[\frac{\partial L}{\partial x_1} \quad \frac{\partial L}{\partial x_2}\right]$$

$$= [g_1 + 2\lambda_1 x_1 \quad g_2 + 2\lambda_1 x_2] = 0$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\frac{1}{2\lambda_1}\underbrace{\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}}_{x(\lambda_1)}$$

$$\Rightarrow x(\lambda_1) = -\frac{1}{2\lambda_1} g$$

② find $\Lambda = \{\lambda_1 : h_1[x(\lambda_1)] = 0\}$

$$h_1[x(\lambda_1)] = x_1(\lambda_1)^2 + x_2(\lambda_1)^2 - 4$$

$$= \|x(\lambda_1)\|^2 - 4 = \frac{1}{4\lambda_1^2}\|g\|^2 - 4$$

$$= 0 \quad \text{iff} \quad \lambda_1 = \pm \frac{\|g\|}{4}$$

so

$$\Lambda = \{\underbrace{-\tfrac{1}{4}\|g\|}_{\bar{\lambda}_1}, \underbrace{+\tfrac{1}{4}\|g\|}_{\tilde{\lambda}_1}\}$$

③ find $\tilde{\Lambda} = \{\lambda_1 \in \Lambda : L[x(\lambda_1),\lambda_1]_{xx} > 0\}$

$$L[x,\lambda_1] = g^T x + \lambda_1(x_1^2 + x_2^2 - 4)$$

$$\Rightarrow L[x,\lambda_1]_{xx} = \begin{bmatrix} 2\lambda_1 & 0 \\ 0 & 2\lambda_1 \end{bmatrix}$$

Therefore:

- $L[x(\bar{\lambda}_1),\bar{\lambda}_1]_{xx} = -\frac{1}{2}\|g\|\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \not> 0$ (#)

  $(\Rightarrow \bar{\lambda}_1 \notin \tilde{\Lambda})$

- $L[x(\tilde{\lambda}_1),\tilde{\lambda}_1]_{xx} = \frac{1}{2}\|g\|\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} > 0$ (#)

  $(\Rightarrow \tilde{\lambda}_1 \in \tilde{\Lambda})$

Hence:

$$\tilde{\Lambda} = \{\tilde{\lambda}_1\} = \{\tfrac{1}{4}\|g\|\}.$$

---

a diagonal matrix is positive definite iff all its diagonal entries are >0.

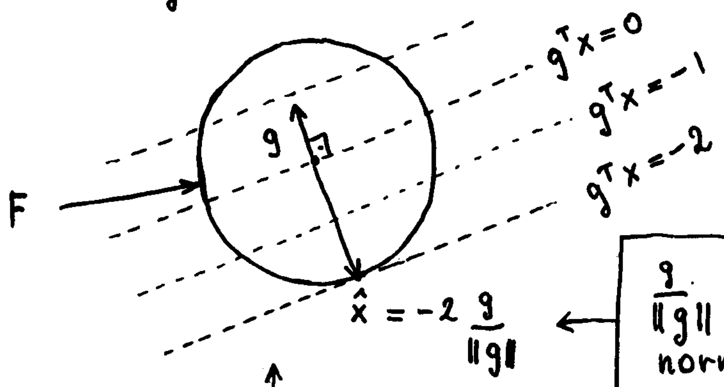
④ find $\tilde{X} = \{x(\lambda_1) : \lambda_1 \in \tilde{\Lambda}\}$

$$= \{x(\tilde{\lambda}_1) = \{-2\frac{g}{\|g\|}\}$$

$=$ our set of local minimizers

So: $\hat{x} = -\frac{2g}{\|g\|}$

For this very simple case, we can check geometrically:



$\frac{g}{\|g\|}$ is the normalized vector g

we see from this diagram that our local minimizer $\hat{x}$ is actually the global minimizer ← generally it is not always so nice.

Equality constrained optimization via duality

In duality theory, the original opt. problem is called the primal problem (P):

Our primal problem is:

$$\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = 0\} \qquad (P)$$

$h : \mathbb{R}^n \to \mathbb{R}^m$

We shall derive an "equivalent" dual problem (D), potentially easier to solve than (P), from which the solution to (P) can be found easily.

Solution of (P) will be aided by study of

$$\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = y\}$$

$\in \mathbb{R}^n$

a generalized problem owing to y

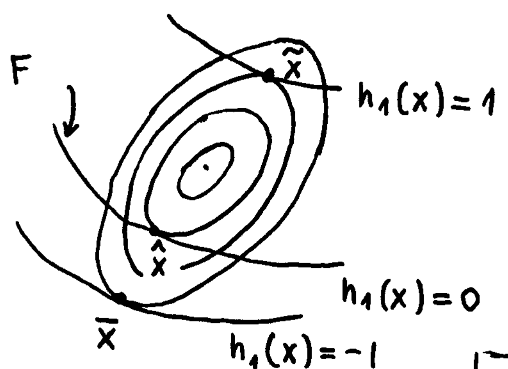
Define :

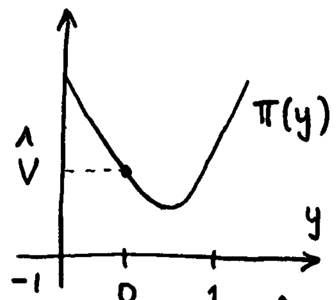
- $\underbrace{\pi(y)}_{\text{the perturbation function}} \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : \underbrace{h(x)=y}_{\text{perturbation of primal constraint } h(x)=0} \}$

- $\hat{V} \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : h(x) = 0 \}$

  = the minimal cost for (P).

(14.1) Th $\quad \pi(0) = \hat{V} \leftarrow$ $\boxed{\text{obvious}}$





$\pi(-1) = V(\bar{x})$
$\pi(0) = V(\hat{x}) = \hat{V}$
$\pi(1) = V(\tilde{x})$

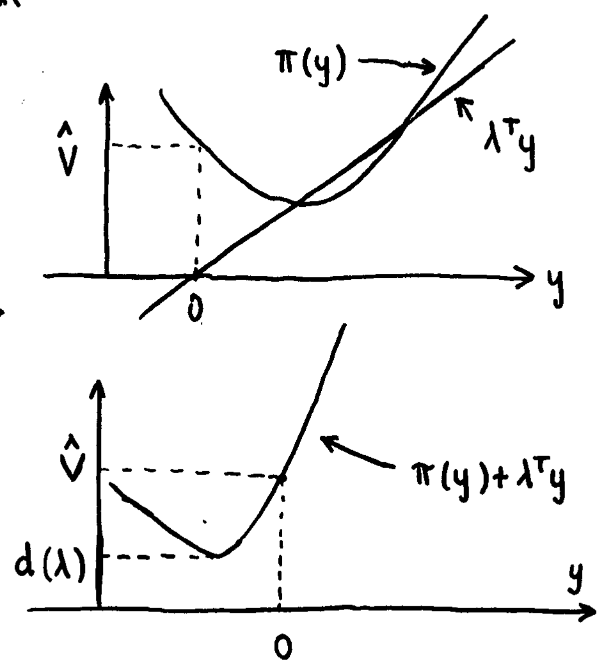
this kind of plot is the key to duality theory

Duality theory relies heavily on the dual function:

$$\underbrace{d(\lambda)}_{\in \mathbb{R}^m} \triangleq \min_{y \in \mathbb{R}^m} [\pi(y) + \lambda^T y] \qquad \text{(DF)}$$

graphically :

adding $\lambda^T y$ "rotates" π about $y=0$



(14.2) Th

(i) $\pi(y) + \lambda^T y \,|_{y=0} = \hat{V} \;, \quad \forall \lambda \in \mathbb{R}^m$

(ii) $d(\lambda) \le \hat{V} \;, \quad \forall \lambda \in \mathbb{R}^m$

$\boxed{\text{obvious}}$

Proof of Th (14.2)(ii):

$$d(\lambda) = \min_{y \in \mathbb{R}^m} [\pi(y) + \lambda^T y]$$

$$\leq [\pi(y) + \lambda^T y]\big|_{y=0}$$

$$= \pi(0) = \hat{V} \quad , \quad \forall \lambda \in \mathbb{R}^m$$

QED.

Evaluation of $d(\lambda)$:

$$d(\lambda) = \min_{y \in \mathbb{R}^m} [\underbrace{\pi(y)}_{\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = y\}} + \lambda^T y]$$

$$\Downarrow$$

$\boxed{\text{finding } d(\lambda) \text{ this way is at least as hard as solving (P) so, this way seems useless as an aid to solving (P)}}$

BUT ... saved by $\downarrow$

(14.3) **Th** $d(\lambda)$ is also given by

$$d(\lambda) = \min_{x \in \mathbb{R}^n} [V(x) + \lambda^T h(x)]$$

$\boxed{\text{unconstrained minimization relatively easy to do using secant alg. etc.}}$

Proof:

(14.3) Let $\mathcal{J}(x,y) \triangleq \begin{cases} V(x) & \text{if } h(x) = y \\ +\infty & \text{if } h(x) \neq y \end{cases}$

$$\theta \triangleq \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \{\mathcal{J}(x,y) + \lambda^T y\}$$

Then:

$$\theta = \min_{x \in \mathbb{R}^n} \Big[\min_{y \in \mathbb{R}^m} \{\mathcal{J}(x,y) + \lambda^T y\}\Big] \qquad (\#)$$

$$= \min_{y \in \mathbb{R}^m} \Big[\min_{x \in \mathbb{R}^n} \{\mathcal{J}(x,y) + \lambda^T y\}\Big] \qquad (£)$$

From (#):

$$\theta = \min_{x\in\mathbb{R}^n}\Big[\min_{y\in\mathbb{R}^m}\{J(x,y)+\lambda^T y\}\Big]$$

$$= \min_{x\in\mathbb{R}^n}\Big[\, J(x,\hat{y})+\lambda^T\hat{y}\,\Big|_{\hat{y}=h(x)}\Big]$$

as $J(x,y)+\lambda^T y = \infty$ if $h(x)\neq y$

$$= \min_{x\in\mathbb{R}^n}\{J(x,h(x))+\lambda^T h(x)\}$$

(14.3)

$$= \min_{x\in\mathbb{R}^n}\{V(x)+\lambda^T h(x)\}$$

$= d(\lambda)$ of Th (14.3)

From (£):

$$\theta = \min_{y\in\mathbb{R}^m}\Big[\min_{x\in\mathbb{R}^n}\{J(x,y)+\lambda^T y\}\Big]$$

$$\Rightarrow \theta = \min_{y\in\mathbb{R}^m}\Big[\min_{x\in\{x\in\mathbb{R}^n:\,h(x)=y\}}\{J(x,y)+\lambda^T y\}\Big]$$

since $J(x,y)+\lambda^T y = +\infty$ if $h(x)\neq y$

$$= \min_{y\in\mathbb{R}^m}\Big[\min_{x\in\{x\in\mathbb{R}^n:\,h(x)=y\}}\{J(x,h(x))+\lambda^T y\}\Big]$$

$$= \min_{y\in\mathbb{R}^m}\Big[\min_{x\in\{x\in\mathbb{R}^n:\,h(x)=y\}}\{V(x)+\lambda^T y\}\Big]$$

$$= \min_{y\in\mathbb{R}^m}\Big[\underbrace{\min_{x\in\{x\in\mathbb{R}^n:\,h(x)=y\}}\{V(x)\}}_{\pi(y)} + \lambda^T y$$

$$= \min_{y\in\mathbb{R}^m}\{\pi(y)+\lambda^T y\}$$

$= d(\lambda)$ of Definition (DF)

Hence:

$$\min_{x \in \mathbb{R}^n} \{V(x) + \lambda^T h(x)\} \overset{(\#)}{=} \Theta \overset{(£)}{=} \min_{y \in \mathbb{R}^m} \{\pi(y) + \lambda^T y\}$$

$$\|\qquad\qquad\qquad\qquad\|$$

$d(\lambda)$ of Th 14.3 $=$ $d(\lambda)$ of Def^n (DF)

same thing.

QED

$\boxed{\text{expression for } d(\lambda) \text{ which is easy to evaluate}}$ $\to$ $d(\lambda)$ of Th 14.3

$\boxed{\text{expression for } d(\lambda) \text{ which has conceptual importance in constrained optimization}}$ $\to$ $d(\lambda)$ of Def^n (DF)

(15.1) <u>Definition</u> (P) is called convex if :

V = convex

$h(x) = Dx - f \leftarrow \boxed{\text{i.e. } h \text{ is affine}}$

(15.2) <u>Th.</u> [(P) is convex]

$\Rightarrow$ [π = convex on $\mathbb{R}^m$] ($)

(15.3) <u>Th</u> (slight extension of Th. (13.3))

If : $f : \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{R}^n$ and

f is differentiable at x :

Then :

$$f(z) \geqslant f(x) + \nabla f(x)^T [z - x] \quad , \forall z \in \mathbb{R}^n$$

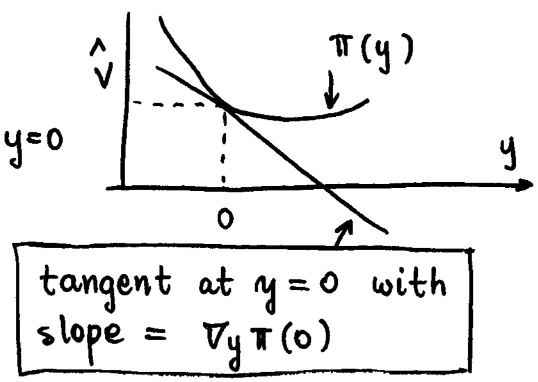
i.e. f lies above its tangent at x.

($) so the situation with π = convex happens for at least one important type of problem

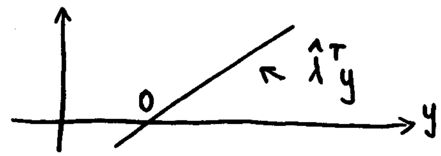
Suppose π is

(i) convex on $\mathbb{R}^m$

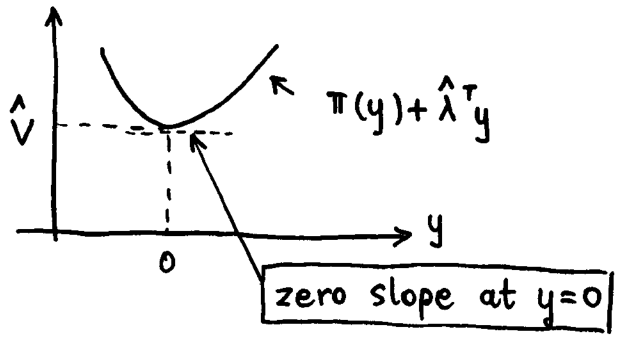
(ii) differentiable at $y=0$



Let $\hat{\lambda} = -\nabla_y \pi(0)$



Then



So :

$$d(\hat{\lambda}) = \min_{y \in \mathbb{R}^m} \pi(y) + \hat{\lambda}^T y \equiv \hat{V} \quad !$$

Hence, we have seen graphically that:

(15.4) **Th** If:

$$\left.\begin{array}{ll}(i) & \pi \text{ is convex on } \mathbb{R}^m \\ (ii) & \pi \text{ is differentiable at } y=0 \\ (iii) & \hat{\lambda} = -\nabla_y \pi(0)\end{array}\right\} \boxed{(H)}$$

Then: $d(\hat{\lambda}) = \hat{V}$.

$\Downarrow$

So, if $(H) = \text{true}$, then

$d(\hat{\lambda}) = \hat{V}$ for some $\hat{\lambda} \in \mathbb{R}^n$

e.g. for $-\nabla_y \pi(0)$

Also,

$d(\lambda) \leq \hat{V}$, $\forall \lambda \in \mathbb{R}^m$

Th (14.2)

Hence:

$$\boxed{\max_{\lambda \in \mathbb{R}^m} d(\lambda) = \hat{V}} \leftarrow \boxed{\text{Problem } (D)}$$

$$\hat{V} = \min_{x \in \mathbb{R}^n} \{ V(x) : h(x) = 0 \}$$

constrained optimization
$\parallel\parallel$
THE PRIMAL PROBLEM (P)

$$\max_{\lambda \in \mathbb{R}^m} d(\lambda) \leftarrow \boxed{\text{unconstrained optimization}} \leftarrow \boxed{\text{THE DUAL PROBLEM}}$$

we know how to solve it
e.g. $\max d = -\min(-d)$

So: we can find $\hat{V}$ for (P) by unconstrained maximization of (D)

OK, but what we really want is a global minimizer $\hat{x}$ for (P).

(15.5) **Th** If:

$$\left.\begin{array}{ll}(i) & \pi \text{ is convex on } \mathbb{R}^m \\ (ii) & \pi \text{ is differentiable at } y=0 \\ (iii) & \hat{\lambda} \in \arg\max\limits_{\lambda \in \mathbb{R}^m} d(\lambda)\end{array}\right\} \text{— (H)}$$

Then:

$$\arg\min (P) = B \cap F \qquad (£)$$

$$\| \triangleright$$

$$\arg\min_{x \in \mathbb{R}^n} \{ V(x) : h(x) = 0 \}$$

where:

$$B = \arg\min_{x \in \mathbb{R}^n} \{ V(x) + \hat{\lambda}^T h(x) \}$$

$$F = \{ x \in \mathbb{R}^n : h(x) = 0 \}$$

(£) This provides a useful mechanism for finding the set of global minimizers for the primal problem from the set of minimizers B, associated with the dual problem. A singleton assumption yields a big simplification.

$\downarrow$

Remark:

__Suppose__: each set arg min contains exactly one element i.e. is a singleton $\boxed{\text{often true}}$

$\Downarrow$

$$\underbrace{\arg\min (P)}_{= \{\hat{x}\}} = \underbrace{B \cap F}_{= \{\tilde{x}\}}$$

Then either: $\tilde{x} \notin F$ or $\underbrace{\tilde{x} \in F}_{\boxed{\text{must be true}}}$

$\Downarrow$

$$\underbrace{\{\hat{x}\} = \{\tilde{x}\} \cap F = \emptyset}_{\boxed{\text{a contradiction}}}$$

So: $\{\hat{x}\} = \{\tilde{x}\} \cap F = \{\tilde{x}\} \Rightarrow \hat{x} = \tilde{x}$.

(15.6) Duality-based procedure for solving (P)

Assume: • π = convex on $\mathbb{R}^m$ ← OK if (P) is convex

• π = differentiable at $y = 0$

not easy to check but usually OK

• that the singleton hypothesis is valid.

often is OK

Find:

$$\{\hat{\lambda}\} = \arg\max_{\lambda \in \mathbb{R}^m} d(\lambda)$$

← unconstrained optimization algs. used

$$\{\tilde{x}\} = \arg\min_{x \in \mathbb{R}^n} \{V(x) + \hat{\lambda}^T h(x)\}$$

unconstrained optimization

Then: $\hat{x} = \tilde{x}$

for primal problem (P)

In practice:

except for V = quadratic & h = affine using procedure (15.6) is harder than it seems since we cannot get a nice (*) formula for $d(\lambda)$

$\Rightarrow$ cannot easily find $\hat{\lambda}$ in this way

Use of a "trick" largely overcomes this problem, and also often makes things work for non-convex (P) as well.

The trick is to modify π to π_r $(r > 0)$:

$$\pi_r(y) \triangleq \pi(y) + \frac{r}{2}\|y\|^2 \leftarrow \boxed{\text{augmented perturbation function}}$$

---

* By a nice formula for $d(\lambda)$ we mean one which can be maximized analytically.
Our formula for $d(\lambda)$ yields a value for $d(\lambda)$ relatively easily but is not easy to maximize analytically.

Then, for big enough r

$$\underset{\uparrow}{\pi_r \cong} \text{ approximately quadratic in } y$$

$\boxed{\text{with nice computational consequences}}$

Substituting π by π_r, we can re-do the "duality theory" using:

$$d_r(\lambda) \triangleq \min_{\lambda \in \mathbb{R}^m} \{ \pi_r(y) + \lambda^T y \} \leftarrow \boxed{\text{augmented dual function}}$$

for which the easily-evaluatable formula is:

$$d_r(\lambda) = \min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \lambda^T h(x) \right\}$$

Then, we get:

$\boxed{\text{still}}$

$$\hat{V} = \max_{\lambda \in \mathbb{R}^m} d_r(\lambda) \quad \text{with} \quad \hat{\lambda} = -\nabla\pi(0)$$

$\boxed{\text{independent of } r}$

and:

$$\hat{x} = \tilde{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \hat{\lambda}^T h(x) \right\}$$

for primal problem
(P)

Nothing much has changed so far ...
so what is the point of making the augmentations?

The important thing is that:

which justifies using Π_r

$$\hat{\lambda} \cong \lambda + r\,h(\bar{x}), \quad \forall \lambda \qquad (\$)$$

depends on λ

with approximation error decreasing as r increases

where:

$$\bar{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \lambda^T h(x) \right\}$$

<u>Amazing consequence:</u>

($) provides a way to find $\hat{\lambda}$ accurately (if r is big enough) <u>without</u> us needing a nice formula for $d(\lambda)$

(or for $d_r(\lambda)$).

<u>Motivation for ($):</u>

(making the "singleton assumption" and using the approach employed in the proof of Th (14.3))

Consider:

$$\min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \left\{ \mathcal{J}(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (\#)$$

$$\mathcal{J}(x,y) = \begin{cases} V(x) & \text{if } h(x) = y \\ +\infty & \text{if } h(x) \neq y \end{cases}$$

Call the minimizers $\bar{x}$ & $\bar{y}$!

$$= \min_{x \in \mathbb{R}^n} \min_{y \in \mathbb{R}^m} \left\{ \mathcal{J}(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

for each x, the minimizing y is $\hat{y}(x) = h(x)$ (£) as in the proof of Th. (14.3)

$= \dots$

$$\dots = \min_{x\in\mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \lambda^T h(x) \right\} \qquad (¥)$$

the minimizer here must be the minimizer $\bar{x}$ for (#) (and vice versa). So, from (£), the minimizer $\bar{y}$ for (#) must be $\bar{y} = \hat{y}(\bar{x}) = h(\bar{x})$.

Summarizing:

- the minimizer $\bar{x}$ for (#) $\equiv$
  $\equiv$ the minimizer x for (¥)
- the minimizer $\bar{y}$ for (#) $\equiv$
  $\equiv h(\bar{x})$ where $\bar{x}$ can be found by minimizing (¥).

Also, going back to (#) again:

$$\min_{\substack{x\in\mathbb{R}^n\\ y\in\mathbb{R}^m}} \left\{ J(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

$$= \min_{y\in\mathbb{R}^m}\ \min_{x\in\mathbb{R}^n} \left\{ J(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

$$\overset{\boxed{\text{see Th (14.3}}}{=} \min_{y\in\mathbb{R}^m} \left\{ \pi(y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (\mathcal{L})$$

the minimizer here must be the minimizer $\bar{y} = h(\bar{x})$ for (#)

$$\cong \min_{y\in\mathbb{R}^m} \left\{ \underbrace{\pi(0) + \nabla\pi(0)^T(y-0)}_{} + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (\beta)$$

1-st-order expansion for $\pi(y)$ with approximation error swamped by $\frac{r}{2}\|y\|^2$ if r is big enough.

a quadratic in y, for which the minimizing y is $-\frac{1}{r}[\nabla\pi(0)+\lambda]$

Hence:

$$-\frac{1}{r}\left[\nabla\pi(0)+\lambda\right] \cong \bar{y} = h(\bar{x})$$

because we argue that, since opt. problems (α) and (β) are approximately the same, their minimizers $\bar{y} = h(\bar{x})$ and $-\frac{1}{r}[\nabla\pi(0)+\lambda]$ are approximately equal.

That is:

$$-\nabla\pi(0) \cong \lambda + rh(\bar{x})$$

with approximation error decreasing as r increases

i.e. $\hat{\lambda} \cong \lambda + rh(\bar{x})$

depends on λ

with approximation error decreasing as r increases

which is ($) , as required.

Now, we can state "the multiplier algorithm" for solving: $\min_{x\in\mathbb{R}^n} \{ V(x) : h(x) = 0 \}$

and based on duality but does not need a nice formula for $d(\lambda)$.

The Multiplier Algorithm

Choose scalars $r_0 > 0$, $\Psi > 1$, $\lambda_0 \in \mathbb{R}^m$

Set $j=0$

1) [At iteration j] Find

$$\bar{x}_j = \arg\min_{x\in\mathbb{R}^n} \left\{ V(x) + \frac{r_j}{2}\|h(x)\|^2 + \lambda_j^T h(x) \right\}$$

using a standard alg. for unconstrained optimization

2) Set $\lambda_{j+1} := \lambda_j + r_j\, h(\bar{x}_j)$

from ($) : this makes λ_{j+1} an approx. to $\hat{\lambda}$ which improves as j increases since r_j increases with j

3) Set $r_{j+1} = \psi r_j$

to increase r_j

4) Set $j := j+1$ and go to 1)

<u>Remark:</u>

If (P) $\neq$ too nasty

- $\lambda_j \longrightarrow \lambda$ ← owing to ($) and r_j increasing with j
- $\bar{x}_j \longrightarrow \hat{x}$ for (P) ← because $\lambda_j \to \lambda$

Stop iterating when:

(A) $\| h(\bar{x}_j) \| < \varepsilon_1 \Rightarrow$ an adequate approxim. to satisfaction of $h(x) = 0$ has been achieved

(B) $\| \bar{x}_j - \bar{x}_{j-1} \| < \varepsilon_2 \Rightarrow$ no significant change during last iteration

Inequality constrained optimization

$$\min_{x\in\mathbb{R}^n} \{V(x) : h(x) \le 0\} \qquad \text{(ICP)}$$

$h : \mathbb{R}^n \to \mathbb{R}^m$

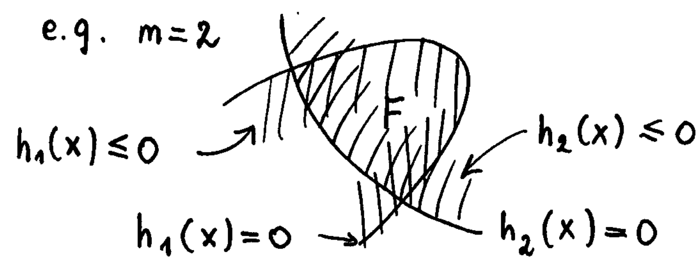
i.e. $h_i(x) \le 0,\ i=1,..,m$

Feasible set $F = \{x\in\mathbb{R}^n : h(x)\le 0\}$

$= \{x\in\mathbb{R}^n : h_i(x)\le 0,\ i=1,..,m\}$

(ICP) is called convex iff $V,\ h_i,\ i=1,..,m$, are convex

e.g. $m=2$



(17.1) Th $\{(\text{ICP}) = \text{convex}\} \Rightarrow \{F = \text{convex}\}$

Optimality on convex F

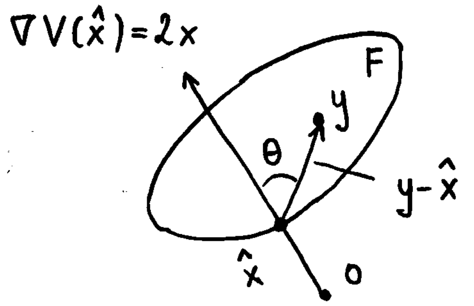
Consider an arbitrary convex set F.

(17.2) Th $\hat{x}\in F$ minimizes a convex, C^1 function V on F, iff:

$$\nabla V(\hat{x})^T(y-\hat{x}) \ge 0, \quad \forall y\in F$$

i.e. $\nabla V(\hat{x})$ and $(y-\hat{x})$ make an angle $\theta \le 90°$ between each other

e.g. $\min_{x\in F} V(x)$ where $V(x) = \|x\|^2$



$\hat{x}$ = point in F nearest to the origin

Remark: (17.2) is a useful optimality condition

Proof of Th. (17.2):

Have to show that:

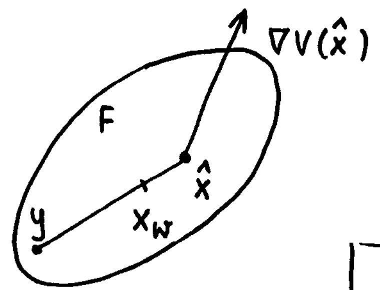
(a) $\hat{x}$ = optimal if $\underbrace{\nabla V(\hat{x})^T (y-\hat{x}) \geq 0,\ \forall y \in F}_{(\alpha)}$

(b) $\hat{x} \neq$ optimal if (α) is false, i.e. if:

$\exists y \in F$ s.t. $\nabla V(\hat{x})^T (y-\hat{x}) < 0$

$\boxed{\text{here we show only (b):}}$

So, consider such a y and x_w:



$$x_w \triangleq \hat{x} + w\,(y-\hat{x}) \underbrace{\in F}_{\boxed{\text{by convexity of } F}},\ \forall w \in [0,1]$$

For small $w > 0$:

$$V(x_w) = V(\hat{x} + \omega(y-\hat{x}))$$
$$\cong V(\hat{x}) + \nabla V(\hat{x})^T [\omega(y-\hat{x})]$$

Hence, for small $\omega > 0$:

$$V(x_w) \cong V(\hat{x}) + \omega \underbrace{\nabla V(\hat{x})^T (y-\hat{x})}_{<0} < V(\hat{x})$$

$\Rightarrow$ for small enough $\omega > 0$: $\begin{cases} x_w \in F \\ V(x_w) < V(\hat{x}) \end{cases}$

$\Rightarrow$ $\hat{x} \neq$ optimal on F

Duality for inequality-constrained optimization

... similar to the equality-constrained case but with a different perturbation function:

$$\pi(y) \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : h(x) \leq y \}$$

(17.3) **Th** $\{(ICP) = \text{convex}\} \Rightarrow$

$\Rightarrow \{\pi = \text{convex on } \mathbb{R}^m\}$

An important property of π for inequalities, NOT shared by π for equalities, is:

(17.4) **Th** $\quad \pi(\bar{y}) \le \pi(\tilde{y})$ if $\tilde{y} \le \bar{y}$

meaning $\tilde{y}_i \le \bar{y}_i, \ \forall i$

Proof: Let $F_y \triangleq \{x \in \mathbb{R}^n : h(x) \le y\}$

Then: if $\tilde{y} \le \bar{y} \Rightarrow F_{\tilde{y}} \subset F_{\bar{y}}$

because:

$$\{x \in F_{\tilde{y}}\} \Rightarrow \{h(x) \le \tilde{y} \le \bar{y}\} \Rightarrow \{x \in F_{\bar{y}}\}$$

Now:

$$\pi(\tilde{y}) = \min_{x \in \mathbb{R}^n} \{V(x) : h(x) \le \tilde{y}\}$$

$$= \min_{x \in F_{\tilde{y}}} V(x) = V(\tilde{x})$$

for some minimizer $\tilde{x} \in F_{\tilde{y}}$

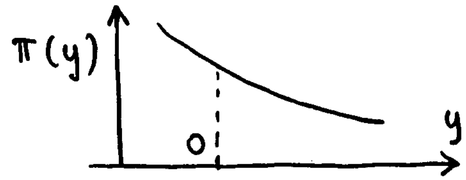
Then if $\tilde{y} \le \bar{y}$

$$\pi(\bar{y}) = \min_{x \in F_{\bar{y}}} V(x) \le V(\tilde{x}) = \pi(\tilde{y})$$

since $\tilde{x} \in F_{\tilde{y}} \subset F_{\bar{y}}$
so $\tilde{x}$ may not be optimal for $\pi(\bar{y})$

i.e. $\pi(\bar{y}) \le \pi(\tilde{y})$ if $\tilde{y} \le \bar{y}$ QED.

Hence $\pi(y)$ decreases as y increases:



(17.5) **Th** If π is differentiable at $y=0$

then: $\nabla \pi(0) \le 0$

meaning $\frac{\partial}{\partial y_i} \pi(0) \le 0, \ i = 1, \dots, m$

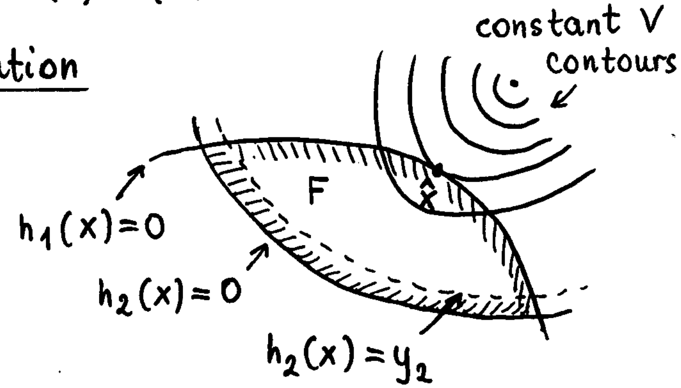
(17.6) **Th** Suppose π is differentiable at $y=0$ and $\hat{x}$ solves (ICP).

Then: $\nabla \pi(0)_i = 0$ if $h_i(\hat{x}) < 0$, $\forall i$

and $\nabla \pi(0)^T h(\hat{x}) = 0$

Plausibilification

case $m = 2$



Here:

$h_1(\hat{x}) = 0$

$h_2(\hat{x}) < 0$ and

$$\underbrace{\min_{x \in \mathbb{R}^2} \left\{ V(x) : \begin{array}{l} h_1(x) \leq 0 \\ h_2(x) \leq y_2 \end{array} \right\}} = V(\hat{x}), \ \forall \text{ small } y_2$$

$$\boxed{\text{i.e. } \pi\left(\begin{bmatrix} 0 \\ y_2 \end{bmatrix}\right) = V(\hat{x}), \ \forall \text{ small } y_2}$$

i.e. $\frac{\partial}{\partial y_2} \pi(0) = 0$ if $h_2(\hat{x}) < 0$

as claimed.

Further:

$$\nabla \pi(0)^T h(\hat{x}) = \frac{\partial \pi(0)}{\partial y_1} \underbrace{h_1(\hat{x})}_{\boxed{=0}} + \underbrace{\frac{\partial \pi}{\partial y_2}(0)}_{\boxed{=0}} \cdot h_2(\hat{x})$$

$= 0$ as claimed

this always happens in the inequality constrained case

For the inequality constrained case, the equality constrained duality theory applies provided the constraint $\lambda \geqslant 0$ is applied.

Exactly as before, define:

(17.7) $$d(\lambda) \triangleq \min_{y \in \mathbb{R}^m} \{ \pi(y) + \lambda^T y \}$$

Then we have the same easily-evaluatable formula for $d(\lambda)$ when $\lambda \geqslant 0$.

(17.8) **Th** If $\lambda \geq 0$: $d(\lambda)$ is also given by:

$$d(\lambda) = \min_{x \in \mathbb{R}^n} \{V(x) + \lambda^T h(x)\}$$

Proof: (some of the proof is a little different).

Let
$$\mathcal{J}(x,y) = \begin{cases} V(x) & \text{if } h(x) \leq y \\ +\infty & \text{if } h(x) \not\leq y \end{cases}$$

Consider
$$\theta = \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \{\mathcal{J}(x,y) + \lambda^T y\}$$

Then:

$$\theta = \min_{x \in \mathbb{R}^n} \underbrace{\min_{y \in \mathbb{R}^m} \{\mathcal{J}(x,y) + \lambda^T y\}}_{= \min_{y \in \{y \in \mathbb{R}^m : h(x) \leq y\}} \{\mathcal{J}(x,y) + \lambda^T y\}}$$

$\boxed{\text{since } \mathcal{J}(x,y) = \infty \text{ if } h(x) \not\leq y}$

Hence:

$$\theta = \min_{x \in \mathbb{R}^n} \min_{y \in \{y \in \mathbb{R}^m : h(x) \leq y\}} \{V(x) + \lambda^T y\}$$

$\boxed{\text{because } \mathcal{J}(x,y) = V(x) \text{ if } h(x) \leq y}$

$$= \min_{x \in \mathbb{R}^n} \min_{y \in \{y \in \mathbb{R}^m : h_i(x) \leq y_i,\ i = 1,..,m\}} \{V(x) + \sum_{i=1}^m \overbrace{\lambda_i}^{\boxed{\geq 0}} y_i\}$$

$$= \min_{x \in \mathbb{R}^n} \{V(x) + \sum_{i=1}^m \lambda_i h_i(x)\}$$

$\boxed{\text{since: } \hat{y}_i = h_i(x) \text{ as } \lambda_i \geq 0}$

$$\equiv \min_{x \in \mathbb{R}^n} \{V(x) + \lambda^T h(x)\}$$

$$\equiv d(\lambda) \text{ of Th (17.8).}$$

But also:

$$\theta = \min_{y \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \{\mathcal{J}(x,y) + \lambda^T y\}$$

So :

$$\theta = \min_{y \in \mathbb{R}^m} \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} \{ J(x,y) + \lambda^T y \}$$

$\boxed{\text{as } J(x,y) = +\infty \text{ if } h(x) \nleq y}$

$$= \min_{y \in \mathbb{R}^m} \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} \{ V(x) + \lambda^T y \}$$

$\boxed{\text{because } J(x,y) = V(x) \text{ if } h(x) \leq y}$

$$= \min_{y \in \mathbb{R}^m} \left\{ \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} V(x) \right\} + \lambda^T y$$

$$= \min_{y \in \mathbb{R}^m} \pi(y) + \lambda^T y = d(\lambda)$$

$\boxed{\text{of Def}^n \text{ (17.7)}}$

Hence : $d(\lambda)$ of Def^n (17.7) $= \theta$

$= d(\lambda)$ of Th. (17.8)

QED

In summary:

Inequality constrained problem:

$$\min_{x\in\mathbb{R}^n} \{ V(x) : h(x) \le 0 \} = \hat{V} \qquad \text{(ICP)}$$

$\boxed{h : \mathbb{R}^n \to \mathbb{R}^m}$

Perturbation function:

$$\pi(y) = \min_{x\in\mathbb{R}^n} \{ V(x) : h(x) \le y \}$$

$$\pi(0) = \hat{V}$$

Dual function:

$$d(\lambda) = \min_{y\in\mathbb{R}^m} \{ \pi(y) + \lambda^T y \} \qquad (18.1)$$

Easily-evaluatable formula for $d(\lambda)$:

$d(\lambda)$ is also given by:

$$d(\lambda) = \min_{x\in\mathbb{R}^n} \{ V(x) + \lambda^T h(x) \} \qquad (18.2)$$

only when $\lambda \geqslant 0$ ← $\boxed{\text{meaning: } \lambda_i \geqslant 0 \ \forall i=1,..m}$

Using definition (18.1) of $d(\lambda)$ until further notice, an outline of duality theory for (ICP) follows:

Exactly as in Th (14.2):

$$d(\lambda) \le \hat{V} \quad , \forall \lambda \in \mathbb{R}^m$$

$$\Downarrow$$

(18.3) $\quad d(\lambda) \le \hat{V} \quad , \forall \lambda \geqslant 0$

Assuming from now on that:

- π is convex on $\mathbb{R}^m$
- π is differentiable at $y=0$

we have:

(18.4) $\quad d(\hat{\lambda}) = \hat{V}$ where $\hat{\lambda} = -\nabla\pi(0) \geqslant 0$

$\boxed{\text{unknown}}$

Hence, from (18.3) and (18.4):

$$(18.5)\quad \underbrace{\max_{\lambda \geq 0} d(\lambda)}_{\boxed{\text{the dual problem for (ICP)}}} = \hat{V} = \underbrace{\min_{x \in \mathbb{R}^n} \{V(x) : h(x) \leq 0\}}_{\boxed{\text{the primal problem (ICP)}}}$$

Since $\lambda \geq 0$ in the dual, we can use the easily-evaluatable formula for d in the dual, making the solution of the dual problem potentially practicable.

We can find $\hat{x}$ for (ICP) from $\hat{\lambda}$:

actually unique, found by maximizing d.

(18.5) $\underline{\underline{\text{Th}}}$ Under the above conditions:

$$\arg\min (\text{ICP})$$
$$\|$$
$$\left\{\arg\min_{x \in \mathbb{R}^n} \{V(x) + \hat{\lambda}^T h(x)\}\right\} \cap \{x \in \mathbb{R}^n : h(x) \leq 0\}$$
$$\|$$
$$\left\{\arg\min_{x \in \mathbb{R}^n} \{V(x) + \hat{\lambda}^T h(x)\} \cap \{x \in \mathbb{R}^n : \underbrace{\hat{\lambda}^T h(x) = 0}\right\}$$

from: $\nabla\pi(0)^T h(\hat{x}) = 0$ $\boxed{\text{Th (17.6)}}$

and $\hat{\lambda} = -\nabla\pi(0)$

If the singleton assumption is valid:

$$\hat{x} = \tilde{x} = \arg\min_{x \in \mathbb{R}^n} \{V(x) + \hat{\lambda}^T h(x)\}.$$

$\boxed{\text{for (ICP)}}$

One can go on to use augmented duals to derive a multiplier alg. for (ICP), all similar to equality constrained problem.

Quadratic programming

(very important for various applications)

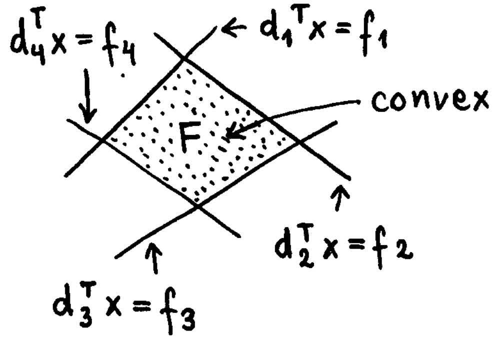
Means of solving of a quadratic program, which is an optimization problem of the kind :

$$\min_{x \in \mathbb{R}^n} \{ \text{standard quadratic} : \underbrace{D}_{\in \mathbb{R}^{m \times n}} x \le f \}$$

Practical applications of QP's include :

- finding search directions for some advanced algs.
- constrained optimal control problems

Feasible set $\quad F = \{ x \in \mathbb{R}^n : D x \le f \}$, $\quad D = \begin{bmatrix} d_1^T \\ \vdots \\ d_m^T \end{bmatrix}$



The simple geometry of such F enables special solution methods to be developed for Q. P. s.

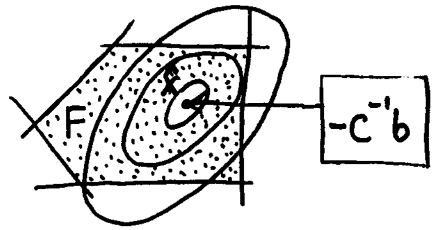
Solution of Q.P.s

For our standard quadratic, the unconstrained global minimizer is $-C^{-1}b$.

For the QP : there are 2 possible situations :

I $\quad -C^{-1}b \in F$

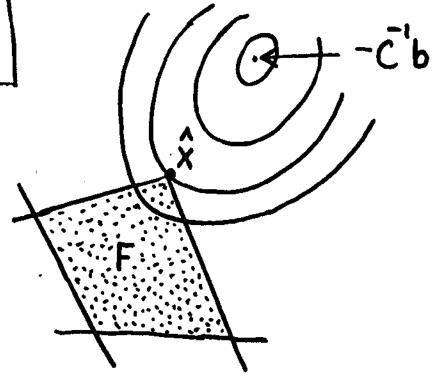
i.e. $D(-C^{-1}b) \le f$



Obviously in this case $\hat{x} = -C^{-1}b$ ... $\boxed{\text{Q.P. solved}}$

II $\quad -C^{-1}b \notin F$ $\quad$ more likely

i.e. $D(-C^{-1}b) \nleq f$



How to find $\hat{x}$?

... Computationally

In case II :

clearly always $\hat{x} \in \underbrace{\text{boundary of } F}_{\left\{x \in \mathbb{R}^n : \text{ some selection of the rows of } D^T x = f \text{ are satisfied as equality constraints and } D^T x \le f \right\}}$

Hence $\hat{x}$ (for Q.P.) can be found by:

- for each possible selection of the rows of $D^T x = f$, find the corresponding optimal x

$\boxed{\text{easy as } V = \text{quadratic and each row of } D^T x = f \text{ is a linear equality constraint}}$

- check for feasibility of any such x

- choose the feasible x which gives the least optimization cost (least among all optimal xs in F).

$\Rightarrow$ optimal $\hat{x}$ (*)

Duality for Q.P.s

The primal Q.P. is:

$$\min_{x \in \mathbb{R}^n} \Big\{ \underbrace{a + b^T x + \tfrac{1}{2} x^T C x}_{\boxed{V(x) = \text{convex}}} : \underbrace{Dx - f \le 0}_{\boxed{h(x) \text{ all } h_i(x) \text{ are convex}}} \Big\}$$

$\Downarrow$

primal problem is convex

$\Downarrow$ $\boxed{(17.3)}$

Π = convex

(*) many Q.P. algs work like this but choose the xs to be tried in an intelligent way so that not all of them are needed.

$$d(\lambda) = \min_{x \in \mathbb{R}^n} \{ V(x) + \lambda^T h(x) \}$$

$$= \min_{x \in \mathbb{R}^n} \{ (a - \lambda^T f) + (b + D^T \lambda)^T x + \tfrac{1}{2} x^T C x \}$$

$$\underset{\boxed{\text{minimum of a standard unconstrained quadratic}}}{=} (a - \lambda^T f) - \tfrac{1}{2} (b + D^T \lambda)^T C^{-1} (b + D^T \lambda)$$

Fact: the duality requirement that:

$\pi(y)$ be differentiable at $y = 0$

can be replaced by the <u>Slater constraint qualification</u>:

$$\exists\, x \in \mathbb{R}^n : \; Dx < f$$

$\boxed{\text{i.e. } (Dx)_i < f_i ,\ \forall i}$

a very reasonable condition,
<u>assumed to be satisfied here</u>

Provided the Slater constraint qualification is satisfied:

$$\underset{\boxed{\text{Q.P. primal}}}{\min_{x \in \mathbb{R}^n} \{ \underset{\boxed{\text{quadratic}}}{V(x)} : Dx \leq f \}} = \hat{V} \underset{(18.5)}{=} \underset{\boxed{\text{quadratic in } \lambda}}{\max_{\substack{\lambda \geq 0 \\ \lambda \in \mathbb{R}^m}} d(\lambda)}$$

$\boxed{\text{the dual Q.P.}}$

$\Downarrow$

$\hat{x}$ for Q.P. can be found easily from
$\hat{\lambda}$ for the dual Q.P. $\leftarrow$ (18.6)

Often : $m \ll n$:

then constrained optimization of the m λ_is can be much easier than direct constrained optimization of the n x_is,

$\Longrightarrow$ hence the potential computational significance of the dual problem

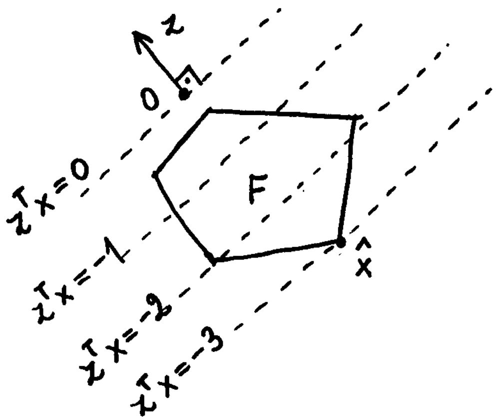
Linear programming

(also very important for applications)

A typical linear program is:

$$\min_{x \in \mathbb{R}^n} \{ z^T x : Dx \le f \}$$

Applications:
- optimization in economics
- optimal control



The corners of F are called its extreme points.

(18.7) **Th** A global minimizer for a L.P. is an extreme point of F.

$$\{ x : Dx \le f \}$$

Consequence:

Many algorithms for solving LPs find all extreme points of F, then find an extreme point giving least cost, and that is an $\hat{x}$. Actually, those algs. iterate so as to consider only extreme points giving lower cost than does the current extreme point.

Duality for L.P.:

The primal problem is:

$$\min \{ \underbrace{z^T x}_{\boxed{V(x) = \text{convex}}} : \underbrace{Dx - f}_{\boxed{h(x),\text{ with convex } h_i}} \le 0 \}$$

$\Downarrow$

primal problem = convex

$\Downarrow$

Π = convex

$$d(\lambda) = \min_{x \in \mathbb{R}^n} \{ V(x) + \lambda^T h(x) \}$$

$$= \min_{x \in \mathbb{R}^n} \{ -\lambda^T f + (z + D^T \lambda)^T x \}$$

(18.8) **Lemma** $d(\lambda) = -\infty$ if $z + D^T\lambda \neq 0$

Proof: If $z + D^T\lambda \neq 0$, consider

$$x \triangleq -\theta(z + D^T\lambda) \quad , \theta\text{-scalar}$$

Then: $-\lambda^T f + (z + D^T\lambda)^T x = -\lambda^T f - \theta \| z + D^T\lambda \|^2$

which can be made as negative as desired by making the scalar θ big enough.

Hence one can make $-\lambda^T f + (z + D^T\lambda)^T x$ as negative as desired by choosing x appropriately, which is what is meant by:

$$\underbrace{\min_{x \in \mathbb{R}^n} \{ -\lambda^T f + (z + D^T\lambda)^T x \}}_{d(\lambda)} = -\infty$$

QED.

Now: assume the Slater constraint qualification is valid.

Then:

$$\min_{x \in \mathbb{R}^n} \{ z^T x : Dx \leq f \} \leftarrow \boxed{\text{the primal LP.}}$$

$$\| $$

$$\hat{V} = \max_{\lambda \geq 0} d(\lambda)$$

since $d(\lambda) = -\infty$ which cannot be maximal if $z + D^T\lambda \neq 0$ we can filter out such non-maximizing λs by adding the constraint $z + D^T\lambda = 0$

$$= \max_{\substack{\lambda \geq 0 \\ z + D^T\lambda = 0}} d(\lambda) = \underbrace{\max_{\substack{\lambda \geq 0 \\ z + D^T\lambda = 0}} (-\lambda^T f)}_{}$$

the dual LP. potentially useful when $m \ll n$

Note: you could not hope to easily write a good alg. for LP or QP. ... use the excellent (and very sophisticated) library programs.