Module – 4.3

MINIMIZATION ALGORITHM

S. Lakshmivarahan

School of Computer Science

University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

MINIMIZATION PROBLEM – 1D

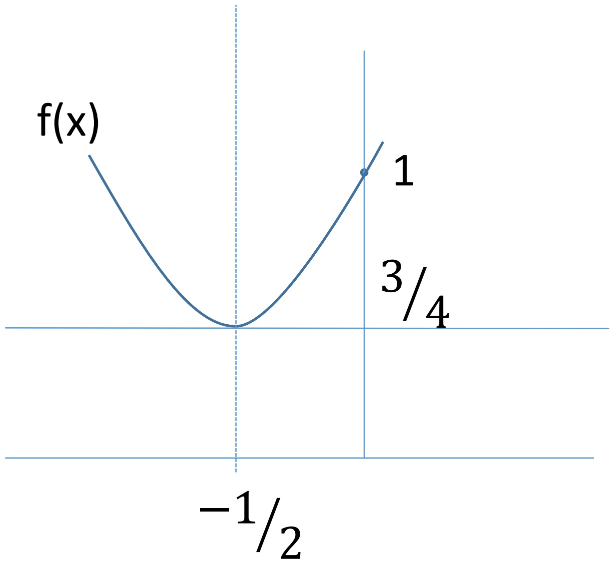
- f: R -> R, be a Convex function
- Example: $f(x) = ax^2 + bx + c$ with $a > 0$
- Rewrite: $f(x) = a[x + \frac{b}{2a}]^2 - (\frac{b^2 - 4ac}{4a})$
- Minimizer $x^* = -\frac{b}{2a}$
- $f(x^*) = -(\frac{b^2 - 4ac}{4a})$
- f(x) is a parabola intersects the x – axis at

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \text{ only if } b^2 > 4ac$$

- Otherwise, f(x) is above the x-axis

MINIMIZATION PROBLEM – 1D

- $f(x) = x^2 + x + 1 = (x + \frac{1}{2})^2 + \frac{3}{4}$
- $x^* = -\frac{1}{2}$ and $f(x^*) = \frac{3}{4}$
- Since $b^2 < 4ac$, $x_{1,2}$ are complex and f(x) lies above the x – axis

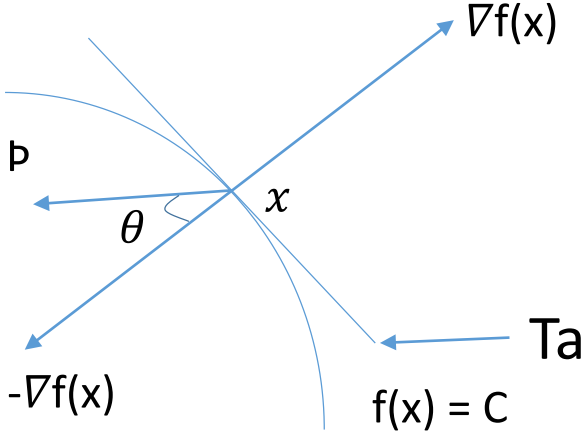


GENERALIZATION – n – DIMENSION

- f:R^n -> R be Convex in R^n
- Example: f(x) = $\frac{1}{2}x^TAx - b^Tx + c$, A – SPD
- ∇f(x) = Ax – b = 0 => x* = A^{-1}b, minimizer of f(x)
- f(x*) = $\frac{2c - b^TA^{-1}b}{2}$, minimum value of f(x)
- Instead of solving Ax = b, we seek to minimize f(x) interatively

A DESCENT DIRECTION

- At any point $x \in R^n$, $\nabla f(x)$ denotes the <u>direction of maximum rate of increase</u>



$|\theta| < 90^0$

$Þ^T[\nabla f(x)] < 0$

Tangent to f(x) at x

- Since $Þ^T[\nabla f(x)] < 0$, Þ is called the <u>descent direction</u>
- f(x) must decrease as we move a small distance along Þ away from x

STEEPEST DESCENT DIRECTION

- Let $\alpha > 0$ be a small real number
- Expand $f(x + \alpha Þ)$ in first order Taylor series

$$f(x + \alpha Þ) \approx f(x) + \alpha Þ^T[\nabla f(x)]$$

$$< f(x) \text{ since } Þ \text{ is a descent direction}$$

- Setting $Þ = - \nabla f(x)$, the steepest descent direction:

$$f(x - \alpha \nabla f(x)) \approx f(x) - \alpha||\nabla f(x)||^2$$

$$< f(x)$$

and we get the maximum rate of decrease in f(x) at x

ROLE OF RESIDUAL VECTOR

- X_k be the current operating point
- Residual $r_k = r(x_k) = -\nabla f(x)$

  $= b - Ax_k$
- R_k is the steepest descent direction of f(x) at x_k
- Since $r_k = 0$ when $x_k = x^* = A^{-1}b$, $\|r_k\|$ is a measure of how far x_k is away from the minimum x^*
- $\|r_k\|$ could be used to test convergence of the iterative minimization

STEEPEST DESCENT FRAMEWORK

- Define the new operating points as

$$x_{k+1} = x_k + \alpha r_k$$

- Then: $f(x_{k+1}) < f(x_k)$ but

$$|f(x_{k+1}) – f(x_k)| \text{ depends on } \alpha$$

- A is called the <u>step length</u> parameter
- At x_k, the direction of search r_k is fixed
- Given x_k and r_k, how to choose α such that we get the maximum decrease in $f(x)$ as we move from x_k to $x_k + \alpha r_k$
- <u>New 1-D minimization problem:</u> minimize g: R -> R where

$$g(\alpha) = f(x_k + \alpha r_k)$$

A DIVIDE AND CONQUER PRINCIPLE

- Given n-dimensional minimization of f(x) is reduced to a sequence of 1-dimensional minimization of g(α) at x_k along the steepest descent direction $r_k = -\nabla f(x_k)$, for k = 0, 1, 2, ….

- This is the basis for the resulting iterative framework for the minimization of f(x)

OPTIMAL STEP LENGTH – QUADRATIC PROBLEM

- Let $f(x) = \frac{1}{2}x^TAx - b^Tx + c$, A – SPD
- Set $x_{k+1} = x_k + \alpha r_k$
- $g(\alpha) = f(x_{k+1}) = f(x_k + \alpha r_k)$

$$= f(x_k) + \frac{1}{2}(r_k^T A r_k)\alpha^2 + (r_k^T A x_k - r_k^T b)\alpha$$

- $g(\alpha)$ is quadratic in α
- Setting: $\frac{dg}{d\alpha} = (r_k^T A r_k)\alpha + r_k^T(Ax_k - b) = 0$
- Minimizer of $g(\alpha)$ is

$$\alpha_k = -\frac{r_k^T(Ax_k - b)}{r_k^T A r_k} = \frac{r_k^T r_k}{r_k^T A r_k} > 0$$

Unless $r_k = 0$

STEEPEST DESCENT/GRADIENT ALGORITHM

- $f(x) = \frac{1}{2}x^TAx - b^Tx + c$, $x_0 \in R^n$ given

$$r_0 = r(x_0) = Ax_0 - b$$

For k = 0, 1, 2, …

<u>Step 1</u> $\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}$ - optimal step length

<u>Step 2</u> $x_{k+1} = x_k + \alpha_k r_k$ – iterates

<u>Step 3</u> Test for convergence. If yes, exit

<u>Step 4</u> $r_{k+1} = r_k - \alpha_k A r_k$ – residual update

ORTHOGONALITY OF RESIDUALS

- Recall that the residual at x_{k+1} is

$$r_{k+1} = b - Ax_{k+1}$$
$$= b - A(x_k + \alpha_k r_k)$$
$$= r_k - \alpha_k A r_k \text{ - The residual update}$$

- Also $r_k^T r_{k+1} = r_k^T(r_k - \alpha_k A r_k)$

$$= r_k^T r_k - \alpha_k r_k^T A r_k$$
$$= 0$$

- That is, $r_{k+1} \perp r_k$
- Convergence question: When is $\lim_{k\to\infty} x_k = x^* = A^{-1}b$?

ERROR AND RESIDUAL VECTORS

- Define the error: $e_k = x_k - x^* = x_k - A^{-1}b$
- Then: $Ae_k = Ax_k - b = -r_k$
- r_k is measurable but e_k is not
- e_k is useful is proving convergence of the sequence $x_0, x_1, x_2, \ldots$
- To prove Convergence: show $\lim_{k\to\infty} e_k = 0$

ENERGY NORM OF THE ERROR e_k

- Define

$$E(x_k) = f(x_k) - f(x^*)$$

- Setting b = Ax* and simplifying

$$E(x_k) = \frac{1}{2}(x_k - x^*)A(x_k - x^*)$$
$$= \frac{1}{2}\, e_k^T A e_k = \frac{1}{2}\, ||e_k||_A^2 > 0$$

unless $e_k = 0$

- $E(x_k)$ is a measure of how far x_k is from x*
- Since A is SPD, $E(x_k) = 0$ if and only if $x_k = x^*$

A FRAME WORK FOR CONVERGENCE PROOF

- Evaluate E(x) along the trajectory and prove that $E(x_k)$ is a decreasing function of k

- Since $E(x_k)$ is bounded below by zero, prove that $E(x_k) \rightarrow 0$ as $k \rightarrow \infty$

- This framework is due to A.Lyapunov and has come to be known as the Lyapunov method

A RECURSIVE RELATION FOR $E(x_k)$

- $E(x_{k+1}) = f(x_{k+1}) – f(x^*)$

- Substituting $x_{k+1} = x_k + \alpha_k r_k$ and simplifying with $b = A^{-1}x^*$, it follows:

$$E(x_{k+1}) = \beta_k E(x_k)$$

$$\beta_k = [\ 1 - \frac{(r_k^T r_k)^2}{(r_k^T A r_k)(r_k^T A^{-1} r_k)}]$$

KANTOROVICH INEQUALITY

- Let $A \in R^n$ be SPD
- Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ be the eigenvalues of A
- Kantorovich inequality states: for any $y \in R^n$

$$\frac{(y^T y)^2}{(y^T A y)(y^T A^{-1} y)} \geq 1 - \left[\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right]^2$$

UPPER BOUND ON β_k: CONDITION NUMBER OF A

- Combining:
- $\beta_k = \left[1 - \frac{(r_k^T r_k)^2}{(r_k^T A r_k)(r_k^T A^{-1} r_k)} \right]$

$$\le \left[\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right]^2 = \left[\frac{\frac{\lambda_1}{\lambda_n} - 1}{\frac{\lambda_1}{\lambda_n} + 1}\right]^2 = \left[\frac{\mathcal{K}_2(A) - 1}{\mathcal{K}_2(A) + 1}\right]^2 = \beta < 1$$

- $\mathcal{K}_2(A) = \frac{\lambda_1}{\lambda_n}$ = condition number of A

  ≥ 1 when A is SPD

CONVERGENCE OF $E(x_k)$

- Hence

$$E(x_{k+1}) \leq \beta E(x_k) \text{ and } \beta < 1$$

- Iterating

$$E(x_k) \leq \beta^k E(x_0) \rightarrow 0 \text{ as } k \rightarrow \infty$$

- Hence, $\lim_{k\rightarrow\infty} \mathrm{E}(\mathrm{x_k})$ = 0 and $\lim_{k\rightarrow\infty} \mathrm{x_k}$ = x*

SUMMARY – MAIN THEOREM

- If $f(x) = \frac{1}{2}x^TAx - b^Tx + c$, and A is SPD then the gradient algorithm, starting from any x_0, Converges to the minimum as $k \to \infty$

- However, the rate if Convergence depends on β which in term depends only on the condition number $\mathcal{K}_2(\mathrm{A})$ of A and not on n, the dimension of the space

ESTIMATION OF THE NUMBER OF ITERATIONS

- For what value of k:

$$\frac{E(x_k)}{E(x_0)} \leq \beta^k = \varepsilon = 10^{-d}$$

- Solving $\beta^k = 10^{-d}$ => $k^* = \left\lceil \frac{d}{\log_{10} \beta^{-1}} \right\rceil$

- That is, for a given β, in k* iterations

$$\frac{E(x_k)}{E(x_0)} \leq 10^{-d}$$

DEPENDENCE OF k* ON β AND $\mathcal{K}_2(\mathrm{A})$

$\mathcal{K}_2(\mathrm{A})$	β	k*
1	0	-
10	0.66942	40
100	0.960788	403
1000	0.996008	4030
10^4	0.9996	40288

EXAMPLE IN R^2

- Consider A = $\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}$, with λ ≥ 1
- f(x) = $\frac{1}{2}x^TAx = \frac{1}{2}(x_1^2 + \lambda_1 x_2^2)$

  $= \frac{x_1^2}{(\sqrt{2})^2} + \frac{x_2^2}{(\sqrt{2\lambda})^2}$
- The minimum of f(α) occurs at $x^* = (0, 0)^T$
- $\nabla f(x_k) = Ax = (x_1, \lambda x_2)^T = -r(x)$
- Set $x_0 = (\lambda, 1)^T$
- Verify $\alpha_0 = \frac{r_0^T r_0}{r_0^T A r_0} = \frac{2}{1+\lambda}$
- $x_1 = x_0 + \alpha_0 r_0 = \frac{\lambda-1}{\lambda+1}\begin{bmatrix} \lambda \\ -1 \end{bmatrix}$

EXAMPLE IN R^2 - CONTINUED

- Continuing

$$x_k = \left(\frac{\lambda-1}{\lambda+1}\right)^k \begin{bmatrix} \lambda \\ (-1)^k \end{bmatrix} \rightarrow 0 \text{ as } k \rightarrow \infty$$

- When λ = 4, $x_k = (0.6)^k \begin{bmatrix} 4 \\ (-1)^k \end{bmatrix}$
- Zig-Zag behavior: Iterates x_0, x_1, x_2, …. exhibit oscillatory behavior which slows the convergence

1-D SEARCH – GENERAL CASE

- Given an operating point x, a descent direction Þ, the optimal step length α is obtained by minimizing

$$g(\alpha) = f(x + \alpha Þ)$$

- Solve

$$\frac{dg}{d\alpha} = [\nabla f(x + \alpha Þ)]^T Þ = 0 \qquad \text{-> (*)}$$

- When f is quadratic => g is quadratic and (*) is linear in α
- When f is not quadratic, (*) can be solved only numerically

QUADRATIC APPROXIMATION TO g(α)

- Compute the following values of g(α):

$$g(0) = f(x)$$

$$g(1) = f(x + Þ)$$

$$\frac{dg(0)}{d\alpha} = [\nabla f(x)]^T Þ$$

- Let $m(\alpha) = a\alpha^2 + b\alpha + c$ be a quadratic approximation to g(α)

QUADRATIC APPROXIMATION TO g(α)

- Set m(0) = g(0) = c

  m(1) = g(1) = a + b + c

  $m'(0) = \left.\frac{dm(\alpha)}{d\alpha}\right|_{\alpha=0} = \left.\frac{dg(\alpha)}{d\alpha}\right|_{\alpha=0} = (2a\alpha + b)|_{\alpha=0} = b$

- Hence a = g(1) – g(0) – g’(0)

  b = g’(0)

  c = g(0)

- Setting $\frac{dm(\alpha)}{d\alpha} = 0$ => optimal step length

  $\alpha = -\frac{b}{2a} = \frac{g\prime(0)}{2[g(1)-g(0)-g'(0)]}$

A LOOK BACK

- Gradient method Converges only asymptotically even for Quadratic functions
- Is finite time Convergence feasible at least theoretically?
- The conjugate direction/conjugate gradient methods can in principle achieve this goal for Quadratic functional

A-CONJUGATE VECTORS

- Let A ∈ R^{nxn} be SPD
- S = {$Þ_0$, $Þ_1$, … $Þ_{n-1}$} be a set of n non-null vectors in R^n
- This set is mutually A-Conjugate if

$$Þ_i^T A Þ_j = 0 \text{ for } i \neq j$$
$$\neq 0 \text{ for } i = j$$

- Extension of the notion of orthogonality
- <u>Claim:</u> if a set S of vectors are A-Conjugate then they are also <u>linearly independent</u>

CONJUGATE VECTORS AS A BASIS FOR R^n

- Let $x_0 \in R^n$ be a fixed vector in R^n
- For any $x \in R^n$:

$$x - x_0 = \alpha_0 Þ_0 + \alpha_1 Þ_1 + \ldots + \alpha_{n-1} Þ_{n-1}$$

- By A-Conjugacy

$$Þ_k^T A(x - x_0) = \sum_{j=0}^{n-1} \alpha_j Þ_k^T A Þ_j = \alpha_k Þ_k^T A Þ_k$$

$$\alpha_k = \frac{Þ_k^T A(x - x_0)}{Þ_k^T A Þ_k}, \quad 0 \le k \le n - 1$$

SOLUTION OF Ax = b USING CONJUGATE VECTORS

- Let $x^* \in R^n$ be the solution of the linear system Ax = b where A is SPD
- Let S = {$Þ_0$, $Þ_1$, … $Þ_{n-1}$} be A – Conjugate
- If x_0 is an initial guess, then

$$A(x^* - x_0) = b - Ax_0 = r_0$$

Is the residual at x_0

- Then

$$x^* = x_0 + \sum_{j=0}^{n-1} \alpha_j Þ_j$$

$$\alpha_k = \frac{Þ_k^T A(x* - x_0)}{Þ_k^T A Þ_k} = \frac{Þ_k^T A r_0}{Þ_k^T A Þ_k}$$

QUADRATIC MINIMIZATION

- Let $A \in R^{nxn}$ be SPD, $b \in R^n$, $c \in R$
- Consider $f(x) = \frac{1}{2}x^TAx - b^Tx + c$
- Minimizer is the solution of $Ax = b$
- Given $Ax = b$, $r(x) = b - Ax$
- Minimize $f(x) = \frac{1}{2}r^T(x)r(x) = \frac{1}{2}(b - Ax)^T(b - Ax)$

$$= \frac{1}{2}b^Tb - b^TAx + \frac{1}{2}x^T(A^TA)x$$

- $\nabla f(x) = (A^TA)x - A^Tb = 0$ => $Ax = b$ if A is SPD

LINEAR TRANSFORMATION – CONJUGATE BASIS

- Define $P = [Þ_0, Þ_1, \dots Þ_{n-1}] \in R^{nxn}$

- $P^TAP = \begin{bmatrix} Þ_0^T \\ Þ_1^T \\ \vdots \\ Þ_{n-1}^T \end{bmatrix} A[Þ_0, Þ_1, \dots Þ_{n-1}]$

  $= Diag(d_0, d_1, \dots d_{n-1}) = D \in R^{nxn}$

  $d_i = Þ_i^T AÞ_i , 0 \le i \le n - 1$

- Let $x = x_0 + P\alpha, \alpha \in R^n$

DECOMPOSITION OF f(x) IN CONJUGATE BASIS

- Define ($r_0 = b - Ax_0$) , $D = P^TAP$

$$G(\alpha) = f(x) = f(x_0 + P\alpha)$$

$$= \frac{1}{2}(x_0 + P\alpha)^T A(x_0 + P\alpha) - b^T(x_0 + P\alpha)$$

$$= (\frac{1}{2} x_0^T A x_o - b^T x_o) + \frac{1}{2} \alpha^T(P^TAP)\alpha - (b - Ax_o)^T p\alpha$$

$$= f(x_0) + \frac{1}{2}\sum_{k=0}^{n-1} \alpha_k^2 d_k - \sum_{k=0}^{n-1} r_0^T Þ_k \alpha_k$$

$$= f(x_0) + \sum_{k=0}^{n-1} g_k(\alpha)$$

$$g_k(\alpha) = \frac{1}{2} d_k \alpha_k^2 - r_0^T Þ_k \alpha_k$$

DIVIDE AND CONQUER

- $\min_{x} f(x) = \min_{\alpha} \mathrm{f}(\mathrm{x}_0 + \alpha Þ)$

$$= \min_{\alpha} G(\alpha)$$

$$= \min_{\alpha}\{\mathrm{f}(\mathrm{x}_0) + \sum_{k=0}^{n-1} \mathrm{g_k}(\alpha_\mathrm{k})\}$$

$$= \sum_{k=0}^{n-1} \min_{\alpha} \mathrm{g_k}(\alpha_\mathrm{k}) \quad (\mathrm{f}(\mathrm{x}_0) \text{ a constant})$$

$$= \text{Minimization n 1-D problems}$$

Since $\mathrm{g_k}(\alpha_\mathrm{k})$ depend only on α_k

1-D MINIMIZATION

- Recall: $g_k(\alpha_k) = \frac{1}{2} d_k \alpha_k^2 - r_0^T Þ_k \alpha_k$
- $\frac{d g_k(\alpha_k)}{d\alpha_k} = d_k \alpha_k - Þ_k^T r_0 = 0$
- Optimal $\alpha_k = \frac{Þ_k^T r_0}{d_k}$

$$= \frac{Þ_k^T (b - Ax_0)}{Þ_k^T A Þk}$$

CONJUGATE DIRECTION – FRAME WORK

- $f(x) = \frac{1}{2}x^TAx - b^Tx$, $x_0 \in R^n$, $r_0 = b - Ax_0$
- Given A-Conjugate set $S = \{Þ_0, Þ_1, \ldots Þ_{n-1}\}$

For k = 0 to n – 1

<u>Step 1:</u> $\alpha_k = \frac{Þ_k^T r_k}{Þ_k^T A Þ_k}$

<u>Step 2:</u> $x_{k+1} = x_k + \alpha_k Þ_k$

<u>Step 3:</u> $r_{k+1} = r_k - \alpha_k A Þ_k$

<u>Step 4:</u> If $r_{k+1} = 0$, then $x^* = x_{k+1}$

VERIFY THE EXPRESSION FOR α_k IN STEP 1

- Given x_k and $Þ_k$
- Consider the 1-D minimization of

$$g(\alpha) = f(x_k + \alpha Þ_k)$$

$$= \frac{1}{2}(x_k + \alpha Þ_k)^T A(x_k + \alpha Þ_k) - b^T(x_k + \alpha Þ_k)$$

$$= f(x_k) + \frac{1}{2}(Þ_k^T A Þ_k)\alpha^2 - (b - Ax_k)^T Þ_k \alpha$$

Minimizer: $\alpha_k = \dfrac{(b - Ax_k)^T Þ_k}{Þ_k^T A Þ_k} = \dfrac{Þ_k^T r_k}{Þ_k^T A Þ_k}$

VERIFY THE EXPRESSION IN STEP 3

- From the step 2:

$$x_{k+1} = x_0 + \alpha_0 Þ_0 + \alpha_1 Þ_1 + \ldots + \alpha_n Þ_n$$

- $r_{k+1} = b - Ax_{k+1}$

$$= b - Ax_0 - \alpha_0 AÞ_0 - \alpha_1 AÞ_1 - \ldots - \alpha_n AÞ_n$$

$$= r_0 - \sum_{j=0}^{k} \alpha_j A Þ_j$$

$$= r_k - \alpha_k AÞ_k$$

RELATIONS BETWEEN r_k AND $Þ_k$

- $Þ_k^T r_{k+1} = Þ_k^T (r_k – \alpha_k A Þ_k)$

  $= 0$ using α_k in step 1
- $r_{k+1} = b – Ax_{k+1} = - \nabla f(x_{k+1})$

$\Rightarrow x_{k+1}$ minimizes f(x) along the line $x_k + \alpha Þ_k$

- Verify

  $Þ_k^T r_{k+1} = Þ_k^T r_{k+2} = \ldots Þ_n^T r_n = 0$

  $Þ_k^T r_k = Þ_k^T r_{k-1} = \ldots Þ_n^T r_0$

EXPANDING SUBSPACE PROPERTY

- From step 2:

$$x_{k+1} = x_0 + \alpha_0 Þ_0 + \alpha_1 Þ_1 + \ldots + \alpha_n Þ_n$$

- $r_{k+1} = b - Ax_{k+1}$

$$= r_0 - \alpha_0 AÞ_0 - \alpha_1 AÞ_1 - \ldots - \alpha_n AÞ_n$$

- Taking inner product with $Þ_j$, $0 \le j \le k - 1$

$$Þ_j^T r_{k+1} = Þ_j^T r_0 - \alpha_j Þ_j^T AÞ_j = 0 \text{ (Step 1)}$$

=> $r_{k+1} \perp \{Þ_0, Þ_1, \ldots Þ_{n-1}\}$

EXPANDING SUBSPACE PROPERTY

- x_{k+1} minimizes f(x) over

  $x \in x_0 + span\{Þ_0, Þ_1, \dots Þ_{n-1}\}$

- x_{k+1} in addition to minimizing along $x_k + \alpha Þ_k$, it also minimizes in the subspace $x_0 + span\{Þ_0, Þ_1, \dots Þ_{n-1}\}$

- Hence x_{n-1} minimizes f(x) in R^n

FINITE TIME CONVERGENCE IN THEORY

- Given f(x) = $\frac{1}{2}x^TAx - b^Tx$,
- An A-Conjugate set s = {$Þ_0$, $Þ_1$, … $Þ_{n-1}$}
- The conjugate direction framework guarantees convergence in at mos n steps
- Implicit assumption: computations are error free

HOW TO FIND A-CONJUGATE SET?

- Given A SPD, consider the eigen-decomposition of A
- $AV_i = V_i\lambda_i \quad 1 \leq i \leq n$
- Let $V = [V_1, V_2, \dots V_n]$

  $\Lambda = \mathrm{Diag}(\lambda_1, \lambda_2, \dots \lambda_n)$
- $AV = V\Lambda$, $VV^T = V^TV = I$
- $V^TAV = \Lambda$ or $A = V\Lambda V^T$

$\Rightarrow$Eigenvectors A are A-Conjugate

- It is computationally demanding to find the complete eigensystem

CONJUGATE GRADIENT (CG) ALGORITHM

- $f(x) = \frac{1}{2}x^T Ax - b^T x$, $x_0 \in R^n$, $r_0 = b - Ax_0$, $Þ_0 = r_0$
- For k = 0 to n – 1

<u>Step 1:</u> $\alpha_k = \frac{Þ_k^T r_k}{Þ_k^T A Þ_k} = \alpha_k = \frac{r_k^T r_k}{Þ_k^T A Þ_k}$

<u>Step 2:</u> $x_{k+1} = x_k + \alpha_k Þ_k$ - <u>Iterates</u>

<u>Step 3:</u> $r_{k+1} = r_k - \alpha_k A Þ_k$ - <u>Residual</u>

<u>Step 4:</u> Test for convergence:

$r_{k+1}^T r_{k+1} < \varepsilon$, exit

<u>Step 5:</u> $\beta_k = - \frac{r_{k+1}^T A Þ_k}{Þ_k^T A Þ_k} = - \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

<u>Step 6:</u> $Þ_{k+1} = r_{k+1} + \beta_k Þ_k$ - <u>Conjugate director</u>

PROPERTIES OF CG ALGORITHM

- The conjugate directions are computed internationally in steps 5 and 6
- Permits alternate choices for α_k and β_k
- $Þ_k$'s are A-Conjugate
- $r_{k+1} \perp r_k$ as in gradient algorithm
- $r_k \perp$ span{$Þ_0$, $Þ_1$, … $Þ_{n-1}$}

PROPERTIES OF CG ALGORITHM

- Span{$Þ_0$, $Þ_1$, … $Þ_{n-1}$}

  = span{r_0, r_1, … r_{n-1}}

  = span{r_0, Ar_0, A^2r_0 … $A^{k-1}r_0$}

  = $KS_k(A, r_0)$ krylov subspace of dimension k generated by A and r_0

CG WITH FINITE PRECISION ARITHMETIC

- Let $x^* = A^{-1}b$ be the optimal solution
- $E_k = x_k - x^*$ - error
- $E(x_k) = \frac{1}{2} e_k^T A e_k = \frac{1}{2} \|e_k\|_A^2$
- With round-off errors, considered as an iterative process
- $\frac{E(xk)}{E(x_0)} \leq 2\left[\frac{\sqrt{\mathcal{K}_2(A)}-1}{\sqrt{\mathcal{K}_2(A)}+1}\right]^k$
- $\mathcal{K}_2(A) = \frac{\lambda_1}{\lambda_n}$, the spectral condition number of A

NUMBER OF ITERATION NEEDED

- Set $2\left[\frac{\sqrt{\mathcal{K}_2(\mathrm{A})}-1}{\sqrt{\mathcal{K}_2(\mathrm{A})}+1}\right]^k \leq \varepsilon = 10^{-d}$
- $k^*\log\left(\frac{\sqrt{\mathcal{K}_2(\mathrm{A})}-1}{\sqrt{\mathcal{K}_2(\mathrm{A})}+1}\right) \leq \log\frac{\varepsilon}{2}$
- $k\left[\log\left(1-\frac{1}{\sqrt{\mathcal{K}_2(\mathrm{A})}}\right)-\log\left(1+\frac{1}{\sqrt{\mathcal{K}_2(\mathrm{A})}}\right)\right] = \log\frac{\varepsilon}{2}$

$\Rightarrow k^* = \frac{\sqrt{\mathcal{K}_2(\mathrm{A})}}{2}\left|\log\frac{\varepsilon}{2}\right| = \frac{(d+1)\sqrt{\mathcal{K}_2(\mathrm{A})}}{2}$

A COMPARISON WITH GRADIENT ALGORITHM

$\varepsilon = 10^{-7}$

$\mathcal{K}_2(\mathrm{A})$	k* (Gradient)	k* (CG)
10	40	24
10^2	403	74
10^3	4030	231
10^4	40288	730

EXERCISES

14.1) Verify that $E(x_{k+1}) = \beta_k E(x_k)$ with $\beta_k = [\ 1 - \frac{(r_k^T r_k)^2}{(r_k^T A r_k)(r_k^T A^{-1} r_k)}]$

14.2) Prove Kantrovich inequality in Slide 17

14.3) Implement the Gradient and Conjugate gradient algorithm in MATLAB

14.4) Let $x = (x_1, x_2)^T$ and $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Consider $f(x) = \frac{1}{2}x^T A x$

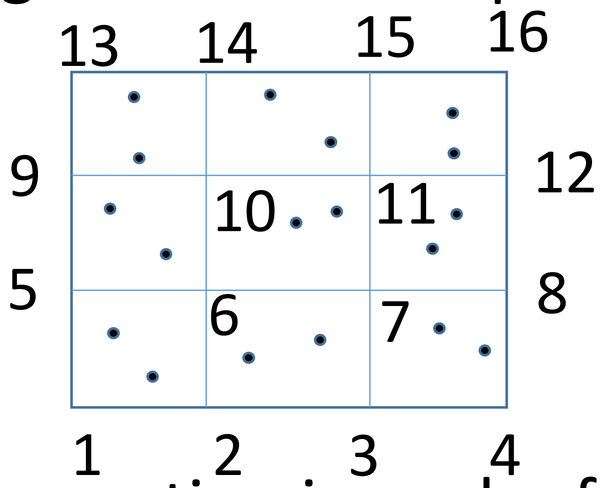
a) Apply the Gradient algorithm and verify that $x_k = (\frac{1}{3})^k \begin{bmatrix} 2 \\ (-1)^k \end{bmatrix}$ with $x_0 = (2, 1)^T$

b) Show that $f(x_{k+1}) = {}^1/_9 f(x_k)$

c) Draw the contour of $f(x)$ and super impose the trajectory $\{x_k\}_{k\geq 0}$ to visually demonstrate convergence

EXERCISES

14.5) Consider a 4x4 grid with $n = 16$ points and q grid boxes as shown



a) Distribute two observation in each of the grid boxes giving a total $m = 18$ observations

b) Build the interpolation matrix $H \in R^{18x16}$

c) Let $Z = (z_1, z_2, \ldots z_{18})^T$ be the observation vector where $z_i = 70 + v_i$, $v_i \sim N(0, \sigma^2)$, $1 \leq i \leq 18$

EXERCISES

d) Construct

$$f(x) = \frac{1}{2}(Z - Hx)^T(Z - Hx)$$

$$= \frac{1}{2}[x^T(H^TH)x - 2Z^THX + Z^TZ]$$

e) Apply the Gradient and Conjugate gradient algorithm to minimize f(x)

f) Plot $f(x_k)$ Vs k for each method comment on your results