$Tr A = \sum_{i=1}^{n} \lambda_i$, $\det A = \prod_{i=1}^{n} \lambda_i$ for $n \times n$ matrix A.

$\vec{x} \in E^n$, $\nabla^2 f(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

Sylvester's Criterion – real, symmetric A positive definite $\Leftrightarrow$ All principal minors of A are positive
$\Updownarrow$ all upper $k \times k$ corners of A are matrices w/ positive determinants.
Also $[a_{11}] < 0 \Rightarrow$ A not positive definite.

Note $\rightarrow f(x) = 0 \in C^k \ \forall k \in \mathbb{N}$.
$\rightarrow g$ Convex $\Leftrightarrow -g$ Concave

Unimodal = single relative minimum.

Min/Max function $f: E^n \rightarrow E$ on $K \subseteq E^n$. $K = E^n$ (unconstrained), $K \neq E^n$ (constrained)

Existence

Soln for: minimize f subject to: $h_i = 0, i = 1, \ldots, m$, $g_j \leq 0$, $j = 1, \ldots, r$

0th order sufficient conditions for rel min

Sufficient condition for solution – f continuous, K compact (closed and bounded) $\Rightarrow f$ has min in K.

$\hookrightarrow h_i, g_j$ continuous $\Rightarrow$ set satisfying $h_i = 0$, $g_j \leq 0$ Closed, $\exists R > 0: K \subseteq B_R(0) = \{x \in E^n : \|x\| < R\} \Rightarrow K$ bounded

$\hookrightarrow$ Unconstrained – $\exists \vec{x}_0 \exists R > 0: \forall x \ \|\vec{x}\| > R \Rightarrow f(\vec{x}_0) < f(\vec{x}) \Rightarrow f$ has min.

for specific x

1st order necessary conditions for rel min

$f \in C^1$ [continuously differentiable] on K, $\vec{x}$ relative min of f in K $\Rightarrow \forall \vec{d}$ feasible directions, $\nabla f(\vec{x}) \cdot \vec{d} \geq 0$

$\vec{x}$ rel min – $\exists \varepsilon > 0 \ \forall \vec{x}': \|\vec{x} - \vec{x}'\| < \varepsilon \Rightarrow f(\vec{x}') \geq f(\vec{x})$, $\vec{d}$ feasible direction at x $\exists \bar{\alpha} > 0 \ \forall 0 < \alpha < \bar{\alpha}: \vec{x} + \alpha \vec{d} \in K$

$\vec{x}$ interior in $K \Rightarrow \forall \vec{d} \ \nabla f(\vec{x}) \cdot \vec{d} \geq 0 \Rightarrow \nabla f(\vec{x}) = \vec{0}$.

for specific x

2nd order necessary conditions for rel min

$\rightarrow f \in C^2$ on K, $\vec{x}$ rel min of $f \Rightarrow \forall \vec{d}$ feasible directions $\nabla f(\vec{x}) \cdot \vec{d} \geq 0$ or $\nabla f(\vec{x}) \cdot \vec{d} = 0 \wedge \vec{d}^T \nabla^2 f(\vec{x}) \vec{d} \geq 0$

$\hookrightarrow \vec{d}^T \nabla^2 f(\vec{x}) \vec{d} \geq 0 \Leftrightarrow$ eigenvalues all ≥ 0 (positive semidefinite)

for specific interior x

2nd order sufficient condition for rel min

$f \in C^2$ on K, $\vec{x}$ interior point of K, $\nabla f(\vec{x}) = 0$, $\nabla^2 f(\vec{x}) > 0$ [$\nabla^2 f(\vec{x})$ positive definite] $\Rightarrow \vec{x}$ strict rel min.

Convex Sets

$C \subseteq E^n$ Convex if $\forall \vec{x}, \vec{y} \in C, \forall 0 < \alpha < 1: \alpha x + (1-\alpha) y \in C$, C, D convex, $\beta \in \mathbb{R} \Rightarrow \beta C$, $C + D$, $\{x\}$, $\emptyset$, $C \cap D$ convex.

Convex Function

f convex on convex set K if $\forall x, y \in K, \forall 0 \leq \alpha \leq 1$ $f(\alpha y + (1-\alpha) x) \leq \alpha f(y) + (1-\alpha) f(x)$

$\hookrightarrow f_1, f_2$ convex, $a \geq 0 \Rightarrow f_1 + f_2$, $a f_1$, $\{x \in K : f(x) \leq a\}$ convex.

$\hookrightarrow f \in C^1$, K convex, f convex $\Leftrightarrow \forall \vec{x}, \vec{y} \in K$, $f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})(\vec{y} - \vec{x})$

$\hookrightarrow f \in C^2$, K convex f convex $\Leftrightarrow \forall \vec{x} \in K$, $\nabla^2 f(\vec{x})$ positive semidefinite.

$\hookrightarrow f$ convex, Epigraph $\{(r, \vec{x}) \mid r > f(\vec{x})\}$ is convex, too.

f convex on convex set $K \Rightarrow$ any local min is a global min.

1st order sufficient condition for global min

$f \in C^1$, convex on convex set K $\exists \vec{x}$ s.t. $\nabla f(\vec{x})(\vec{y} - \vec{x}) \geq 0 \ \forall \vec{y} \in K \Rightarrow \vec{x}$ global min of f on K.

Hyperplane – $(n-1)$ deg subspace, $\{\vec{x} \in E^n : a^T \vec{x} = c\}$ $\vec{a} \neq \vec{0}$, $c \in \mathbb{R}$. For Hyperplane H, $H_+ = \{\vec{x} \in E^n \mid a^T x \geq c\}$ upper half space, $H_- = \{\vec{x} \in E^n \mid a^T x \leq c\}$ lower half space

Separating Hyperplane Theorem

B, C convex, no interior intersection, then there is a hyperplane separating them. If only intersect at boundary, $\exists H$ and $B \cap C \subseteq H$, $B \subseteq H_+$, $C \subseteq H_-$. ie, $H = \{\vec{x} \in E^n \mid a^T x = c\}$ $\forall b \in B, d \in C: a^T b \geq c$, $a^T d \leq c$.

f convex on K convex, $\vec{x}$ min of $f \Rightarrow \exists \vec{\lambda} \neq 0$ s.t. $\vec{x}$ satisfies: (1) min $f(x) + \lambda^T x$, $x \in E^n$ (2) max $\lambda^T x$, $x \in K$.

Convex polytope $\{\vec{x} \in E^n \mid \vec{a}_1 \vec{x} \leq c_1, \ldots, \vec{a}_n \vec{x} \leq c_n\}$, for $\vec{a}_i \in E^n$, $c_i \in \mathbb{R}$. Polyhedron = bounded convex polytope.

Algorithms

Algorithm – $\mathbb{A}: X \rightarrow P(X) = \{$outputs you could get by running some program w/ procedure you have in mind$\} \rightarrow$ Point-Set Map

$\hookrightarrow$ generates: $x_0, x_1 \in \mathbb{A}(x_0), x_2 \in \mathbb{A}(x_1), \ldots = \{x_k\}$. Goal Set – $\Gamma \subseteq X$ (usually $\{$mins of $f\}$)

Descent Function – $Z: X \rightarrow \mathbb{R}$ s.t. 1) $x_k \notin \Gamma, x_{k+1} \in \mathbb{A}(x_k) \Rightarrow Z(x_k) > Z(x_{k+1})$ 2) $x_k \in \Gamma, x_{k+1} \in \mathbb{A}(x_k) \Rightarrow Z(x_k) \geq Z(x_{k+1})$ (Z continuous.)

$\mathbb{A}: X \rightarrow Y$ Closed at $x \in X$ if: $x_k \rightarrow x$, $x_k \in X$, $y_k \rightarrow y$, $y_k \in \mathbb{A}(x_k) \Rightarrow y \in \mathbb{A}(x)$, If $\mathbb{A}$ point to point, continuous $\Rightarrow \mathbb{A}$ closed

$\hookrightarrow \mathbb{A}: X \rightarrow Y$, $\beta: Y \rightarrow Z$ point set maps, $\mathbb{C} = \beta \circ \mathbb{A}: X \rightarrow Z$, $\mathbb{C}(x) = \bigcup_{y \in \mathbb{A}(x)} \beta(y)$, $\mathbb{A}$ closed, β closed on $\mathbb{A}(x) = \bigcup_{x \in X} \mathbb{A}(x)$, Y compact $\Rightarrow \mathbb{C}$ closed.

Global Convergence Theorem – Let $\mathbb{A}: X \rightarrow P(X)$ generate $\{x_k\}$ arbitrarily by $x_{k+1} \in \mathbb{A}(x_k)$, picking $x_0 \in X$. Let $\Gamma \subseteq X$: $\{x_k\}$ compact (X compact or $\{x_k\}$ bounded), $Z: X \rightarrow \mathbb{R}$ descent function for $\mathbb{A}$ and Γ, $\mathbb{A}$ closed $\Rightarrow x_k \rightarrow x \in \Gamma$ (each subsequence of $\{x_k\}$ converges to something in Γ)

Line Search

Newtons Method - $f\in C^2$, x^* min of f, $f'(x^*)=0$, $f''(x^*)\neq 0$ x_0 close to $x^*\Rightarrow\{x_k\}$ produced by Newtons method $\to x^*$

$$x_{k+1}=x_k-\frac{f'(x_k)}{f''(x_k)}$$

False Position - $g\in C^2$, $g(x^*)=0$, $g'(x^*)\neq 0$, x_0 close to $x^*\Rightarrow\{x_k\}$ produced by this $\to x^*$

$$x_{k+1}=x_k-g(x_k)\left[\frac{x_k-x_{k-1}}{g(x_k)-g(x_{k-1})}\right]$$

Quadratic Fit - $F\in C^2$, f unimodal. Find $x_1<x_2<x_3$ w/ $f(x_1), f(x_3)>f(x_2)\Rightarrow x^*\in[x_1,x_3]$

General $S: E^{2n}\to P(E^n)$ $S(\vec{x},\vec{d})=\{\vec{y}\mid \vec{y}=\vec{x}+\alpha\vec{d}$, $\alpha\geq 0$ and $f(\vec{y})=\min_{0\leq\alpha<\infty} f(\vec{x}+\alpha\vec{d})$

↳ f continuous $\Rightarrow S$ closed at points $(\vec{x},\vec{d})$, $\vec{d}\neq\vec{0}$.

Stopping Criterea

Percentage test - Stop when $|\alpha_k-\alpha^*|\leq c\alpha^*$, $f(\vec{x}+\vec{d}\alpha^*)=\min_{0\leq\alpha<\infty}(f(\vec{x}+\alpha\vec{d}))\to$ Need exact min to be on line, but don't need to know α^*!

Armijo, Goldstein, Wolfe - other tests which decide if estimate of α^* too high or low.

Gradient Descent

$A=S\circ G$, $G(\vec{x})=(\vec{x},-g(\vec{x}))$ (g = gradient), $\Gamma=\{\vec{x}\mid\nabla f(\vec{x})=0\}$, $Z(\vec{x})=f(\vec{x})$ a descent function

$\{x_k\}$ bounded $\Rightarrow A$ converges.

Order of Convergence

$\{r_k\}\to r^*$ order $\{r_k\}$ is largest $p\geq 0$ $\overline{\lim_{k\to\infty}}\frac{|r_{k+1}-r^*|}{|r_k-r^*|^p}<\infty$ linear convergence: $p=1$, $\lim_{k\to\infty}\frac{|r_{k+1}-r^*|}{|r_k-r^*|}=\beta<1$ speed of convergence (smaller = better)

$\lim_{k\to\infty}\frac{r_{k+1}}{r_k^p}=c>0\Rightarrow$ order of convergence is p and $r_{k+1}\approx c r_k^p$ (large p, small c = faster convergence).

Convergence of Grad. Descent

for Quadratic $f(\vec{x})=\frac{1}{2}\vec{x}^TQ\vec{x}+\vec{b}^T\vec{x}$, Q pos. def, symmetric. $\exists!$ global min $=Q^{-1}\vec{b}$ order of convergence 1 and $\beta=\frac{(a-A)^2}{(a+A)^2}=\left(\frac{1-r}{1+r}\right)^2$ $A\to$ largest e-value, $a\to$ smallest e-value

r = Condition # $=\frac{A}{a}$ ⊛ ROC unaffected by choice of x_0, only depends on a, A.

Convex Combination - $\alpha_1y_1+\dots+\alpha_ny_n$, $\alpha_i\geq 0$, $\sum_{i=1}^n\alpha_i=1$, $\in$ Convex hull of y_i's.

Non-Quadratic Use Hessian $Q(x)$ of $f(x)$, $\{x_k\}$ (generated by steepest descent) $\to x^*\leftarrow$ relative min, order of convergence $=1$, R.O.C $=\left(\frac{a-A}{a+A}\right)^2$ $A\to$ largest e-value of $Q(x)$, $a\to$ smallest e-value of $Q(x)$

Newtons Method

$x_{k+1}=x_k-(F(x_k))^{-1}\nabla f(x_k)^T$ [Compare w/ 1-var $x_{k+1}=x_k-f''(x_k)^{-1}f'(x_k)$]

$\to$ if $f\in C^3$, x^* rel. min, $F(x^*)$ pos-def, x_0 close to $x^*\Rightarrow\{x_k\}\to x^*$ w/ order of convergence 2.

General Class

$x_{k+1}=x_k-\alpha_kM_k\nabla f(x_k)^T$ Steepest descent: α_k output of S, $M_k=I_k$; N.M.: $\alpha_k=1$, $M_k=F^{-1}(x_k)$

↳ vector determining directional derivative.

$\to$ for far from $x^*\Rightarrow$ large $\varepsilon_k\Rightarrow$ act like grad descent.
$\to$ Near $x^*\Rightarrow$ small $\varepsilon_k\Rightarrow$ act like N.M.

Levenberg-Marquardt

Modified NM, $x_{k+1}=x_k-\alpha_k(\varepsilon_kI+F(x_k))^{-1}\nabla f(x_k)^T$. ($\to M_k\to$ pumps up diagonal) Chose $\delta>0$ and get e-vals of $F(x_k)$, pick ε_k = smallest constant so $\varepsilon_kI+F(x_k)$ has all e-vals $\geq\delta$. Need small δ so that near x^* looks like $F(x_k)$ [N.M.], not too small or e-vals close to 0 and numerically unstable. Need large δ so that far from x^* looks like $\nabla f(x_k)$ [grad descent]

Coordinate Descent

Travel in direction $\vec{e}_i, -\vec{e}_i$ (ith unit vector) at each step. How to chose direction: $\to$ Cyclic: $1,2,\dots,n,1,2,\dots$ $\to$ Double sweep: $1,2,\dots,n,n-1,\dots,2,1,2$ $\to$ Randomly $\to$ Gauss-Southwell: pick $\vec{e}_i$ steepest according to $\nabla f(x_k)$. $\{x_k\}\to x^*$ if $f\in C^1$, $n-1$ iterations = 1 iteration of steepest descent.

Spacer Steps

$\to$ fix algorithm C w/ $Z(C(x))\leq Z(x)$, Z descent function. C may not converge to solution, or too complex to analyze.

Apply another algorithm B which does converge, apply B instead of C every fixed # of steps infinitely often, then this will converge.

↳ $x_k\to x_{k+1}$: $x_{k+1}\in B(x_k)$ instead of $x_{k+1}\in C(x_k)$.

neg def = local max
pos def = local min

Gram-Schmidt Process: begin with $v_1, \dots, v_k$ linearly indp vectors:

Let $u_1 = v_1$ and then $e_j = \frac{v_j}{\|v_j\|}$

$$u_k = v_k - \sum_{j=1}^{k-1} \text{Proj}_{u_j}(v_j) = v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, u_j \rangle}{\|u_j\|^2} u_j$$

Conj Direction Methods

Q pos. def. Symmetric $n\times n$ matrix $u, v \in E^n$, u, v Q-orthogonal $\Leftrightarrow u^TQv = 0$. $\{d_1, \dots, d_k\}$ Q-orthogonal $\Leftrightarrow d_i^TQd_j = 0 \ \forall i \neq j \leq n$

Q orthogonal $\Rightarrow$ linear indp.

- If $f(x) = \frac{1}{2}x^TQx - b^Tx$, $Qx^* = b$, for $\{d_0, \dots, d_{n-1}\}$ Q-orthogonal, $x^* = \sum_{i=0}^{n-1} \alpha_i d_i$, $\alpha_i = \frac{d_i^Tb}{d_i^TQd_i}$
- taking $x_{k+1} = x_k - \alpha_k d_k$, $\alpha_k = \frac{d_k^Tg_k}{d_k^TQd_k}$, $g_k = Qx_k - b$, $\{x_k\} \to x^*$ (w/ $Qx^* = b$) in n steps. $x^* = \sum_{k=0}^{n-1} \frac{d_k^Tg_k}{d_k^TQd_k} d_k$

Conjugate Gradient Method - Start at x_0, let $d_0 = -g_0 = -(Qx_0 - b)$, $x_{k+1} = x_k + \alpha_k d_k$, $\alpha_k = -g^T(x_k)d_k / d_k^TQd_k$

$\hookrightarrow x_n = x^*$, each x_k minimizes f on $x_0 + B_k = \text{span}\{d_0, \dots d_{k-1}\}$. $d_{k+1} = -g_{k+1} + \beta_k x_k$ $g_k = Qx_k - b$ $\beta_k = \frac{g_{k+1}^TQd_k}{d_k^TQd_k}$

* $\forall k$, $\{d_0, \dots, d_k\}$ will be Q-conjugate. $\beta_k = \frac{g_{k+1}^Tg_{k+1}}{g_k^Tg_k}$ ← no Q here!, Converges in n steps for f Quadratic!

Naive - let $g_k = \nabla f(x_k) \to$ expensive, doesn't converge.

Non quadratic case - **Partial CGM** - reset x_0 to x_i every n steps: $d_i = -g(x_i)$

Fletcher Reeves - $d_0 = -g_0$, $x_{k+1} = x_k - \alpha_k d_k$ ($\min_\alpha (x_k + \alpha d_k)$) $d_{k+1} = -g_{k+1} + \beta_k d_k$, $\beta_k = \frac{g_{k+1}^Tg_{k+1}}{g_k^Tg_k}$

Quasi Newton Methods

$x_{k+1} = x_k - \alpha_k S_k \nabla f^T(x_k)$, $S_k \approx F(x)^{-1}$ quadratic - S_k pos. def., Symmetric, rate $= \lim_{k\to\infty} \left(\frac{B_k - b_k}{B_k + b_k}\right)^2$ B_k, b_k max/min e-values of S_kQ.

Approximate F^{-1} **Rank 1 correction** - $H_{k+1} = H_k + (p_k - H_kq_k)(p_k - H_kq_k)^T / q_k^T(p_k - H_kq_k)$, $H_k \to F^{-1}$ in $\leq n$ steps, since $p_i = H_{k+1}q_i$ $i \leq k$ ($H_{k+1} \approx F^{-1}$)

$p_k = x_{k+1} - x_k$, $q_k = g_{k+1} - g_k \Rightarrow q_k \cong Fp_k$. * pos def. only if $q^T(p - H_kq_k) > 0$, not guaranteed! Can have numerical issues.

(rank 2 correction)

DFP - Start w/ symmetric pos def matrix H_0, x_0, ① $d_k = -H_kg_k$ ② $\min_\alpha f(x_k + \alpha d_k) = \alpha_k$, get x_{k+1}, g_{k+1}, let $p_k = \alpha_k d_k$

③ set $q_k = g_{k+1} - g_k$, $H_{k+1} = H_k + \frac{p_kp_k^T}{p_k^Tq} - \frac{H_kq_kq_k^TH_k}{q_k^TH_kq_k}$ * preserves pos def, p_k are F-conjugate (Hessian) $\Rightarrow$ is a conjugate directions method.

Broyden - Satisfy $q_i = Fp_i$ instead of $F^{-1}q_i = p_i$. Generate $B_k \approx F_k$. Update B_{k+1} from B_k using Complimentary formula from H_{k+1} to H_k, swapping p_i with q_i and H_k with B_k. $[B + ab^T]^{-1} = B^{-1} - \frac{B^{-1}ab^TB^{-1}}{1 + b^TB^{-1}a}$

Pure: pick $\varphi \in \mathbb{R}$, $H_{k+1}^{\varphi} = (1-\varphi)H_{k+1}^{DFP} + \varphi H_{k+1}^{BFGS}$ BFGS → Broyden version of DFP, and $\varphi = 0 \Rightarrow$ DFP, $\varphi = 1 \Rightarrow$ BFGS.

Non Pure: Don't pick φ in advance, have a sequence $\{\varphi_i\}$ * $\alpha_k = \min(f_k + \alpha_k d_k) \Rightarrow$ All Broyden methods produce exactly the same x_{k+1}'s!

Lagrange Multipliers

$\min f(\vec{x})$ subject to $\vec{h}(\vec{x}) = (h_i(\vec{x}))_{i=1}^m = 0$, $\vec{g}(\vec{x}) = (g_j(\vec{x}))_{j=1}^p \leq 0$, $\vec{x} \in E^n$, $m \leq n$, $f, g_j, h_i \in C^2$

Active Constraints - restricts feasible directions at x^*, since x^* at boundary of $\{x \mid g_j(x) \leq 0\}$, inactive $\Rightarrow x \in \{x \mid g_j(x) < 0\}$ (interior point)

regular point - x^* w/ $h(x^*) = 0 \wedge \nabla h_1(x^*) \cdots \nabla h_m(x^*)$ L.I.

1st order necc. Conditions $\Rightarrow$ x^* regular, local min $\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$; Solve $n+m$ eqns $h(x^*) = 0$ w.r.t $\underset{m\times 1}{\lambda}, \underset{n\times 1}{x^*}$ Change to definite for **2nd order sufficient Conditions**

2nd order necc. Conditions $\Rightarrow$ x^* regular, local min; x^* solves $n+m$ eqns and $\nabla^2 f(x^*) + \lambda^T \nabla^2 h(x^*)$ positive semi definite on tangent plane of h at x^* $\{y \mid \nabla h(x^*) \cdot y = 0\}$

$\hookrightarrow$ Find orthonormal Basis E of M, $L_M = E^TLE$, check if L_M pos/neg def.

OR $\hookrightarrow$ Solve $p(\lambda) = \det \begin{bmatrix} 0 & \nabla h \\ -\nabla h^T & L - \lambda I \end{bmatrix}$ ($m\times n$ ∇h, $n\times(n-m)$), each λ is an e-value of L_M.

OR $\hookrightarrow$ Bordered Hessian Test → L_M pos def $\Leftrightarrow$ last $n-m$ principal minors of $\begin{array}{cc} \begin{bmatrix} 0 & \nabla h \\ \nabla h^T & L \end{bmatrix} & \begin{matrix} \}m \\ \}n \end{matrix} \\ \underbrace{}_{m}\underbrace{}_{n} & \end{array}$ have sign $(-1)^m$

KKT Conditions

If $\vec{x}^*$ reg point of Constraint set K, which is a local min of f on $K \Rightarrow \exists \vec{\lambda} \in E^m, \exists \vec{\mu} \in E^p$, with $\vec{\mu} \geq 0$ s.t. $\nabla f(\vec{x}^*) + \vec{\lambda}^T \nabla h(\vec{x}^*) + \vec{\mu}^T \nabla g(\vec{x}^*) = 0$, and $\mu^T g(\vec{x}^*) = 0$.

$\boxed{\text{Modified Newtons Method}}$ – Globally Convergent in Finitely Many Steps [Quadratic Case]

– r.o.c = $\lim_{k\to\infty}\left(\frac{B_k - b_k}{B_k + b_k}\right)^2$ B_k, b_k largest, smallest e-values of $M_k \nabla f(x^*)$ [Non quad case]

single step convergence

(order >1) (order 1)
finite > superlinear > linear

General formula: $x_{k+1} = x_k - \alpha_k M_k \nabla f(x_k)^T$

must be +ve definite to guarantee this as a descent method for small α.

Steepest Descent

$A = S \circ G$ (how far in direction to go; direction to go.)

→ α_k output of S, $M_k = I_k$

→ Convergence → Globally Convergent

→ Order 1, R.O.C = $\left(\frac{A-a}{A+a}\right)^2$ [Non Quadratic]

A, a largest + smallest E-vals of $\nabla^2 f(x^*)$

→ $f(x_{k+1}) - f(x^*) \le (1 - \frac{a}{A})(f(x_k) - f^*)$

↳ S.D. makes progress even when far from sol'n!

↳ inexact line search doesn't change the rate much!

Coordinate Descent

→ ignore $\alpha_k M_k \nabla f(x_k)^T$, use $\vec{e}_i$ (or $-\vec{e}_i$)

→ Convergence – slower than S.D. (1 step S.D. = n-1 steps C.D.) – global convergence

→ easy to implement.

Newtons Method

→ $\alpha_k = 1$, $M_k = [\nabla^2 f(x_k)]^{-1}$

→ Convergence – locally convergent

– order 2 (faster than S.D.)

→ Need to ensure M_k positive definite, or

→ $\alpha_k M_k \nabla f(x_k)^T$ may change sign and move away from min! Storage of $\nabla^2 f(x_k)^{-1}$ costly!

→ Sensitive to E-value structure of Hessian, trouble if Hessian hard to compute.

Levenberg-Marquardt

→ Let $M_k = [\varepsilon_k I_k + \nabla^2 f(x_k)^{-1}]$

↳ small $\varepsilon \Rightarrow M_k \approx \nabla^2 f(x_k)^{-1} \Rightarrow$ fast convergence near sol'n since like N.M. near x^*

↳ large $\varepsilon \Rightarrow M_k \approx \varepsilon_k \nabla f(x_k)^T \Rightarrow$ global convergence since like S.D. far from x^*

+ve Def!

Convergence – Global!, order 2 near x^*, contingent on ε.

$M=0$

Partial Conjugate Gradient

→ faster than S.D, less info needed than N.M.

→ Idea: can get d_i from Q-conj d_i w/o x^*, using b, $(Qx^* = b)$

→ more efficient as movement along d_i will not undo progress made in previous iterations ⇒ each step makes more progress than d_i's step of S.D.

→ conjugate gradient procedure carried out for $m+1 < n$ steps, then restarting from current point for $m+1$ more steps taken.

→ n-steps [Quadratic]

$M = n-1$

Quadratic Approximation

Set starting matrix as I_n

→ $g_k \approx \nabla f(x_k)^T$, $Q \approx \nabla^2 f(x_k)$

→ after n-1 steps restart and do pure gradient step.

→ No line search

→ $F(x_k)$ evaluated at each point, impractical! not globally convergent! [Non quadratic]

Line Search Methods

→ use line search in place of computing $\nabla^2 f(x_k)$

→ Globally convergent, as each restart is a step of gradient descent (spacer step)

Costly

→ Convergence linear super, R.O.C. dependent on E-vals of $\nabla^2 f(x^*)$ [Non-quadratic]

Broyden-Fletcher-Goldfarb-Shanno (BFGS)

→ Complimentary formula for DFP, then invert resultant $\nabla^2 f(x)$ to get $\nabla^2 f(x)^{-1}$.

→ Superior performance to DFP.

Quasi-Newton Methods

→ use gradient and that $g(x_{k+1}) - g(x_k)\nabla^2 f(x_k)^{-1} \cong (x_{k+1} - x_k)$ (*)

and $\nabla^2 f(x_n)^{-1} \approx Q^{-1}$ [equality if f Q(x) ⇒ $\nabla^2 f(x^*) = Q$]

Then use Newtons Method w/ this approx.

Convergence → global if restarted every n steps, as each restart begins w/ a S.D. step (spacer step)

↳ superlinear as approximates C.D.M, but takes n steps to get from x_k to x_{k+1}. (Slow for large n)

↳ Sensitive to E-values, inaccurate line searches.

Rank One Correction

→ update $\nabla^2 f(x_j)$ w/ rank 1 matrix maintaining symmetry and (*)

→ May not preserve +ve def of $\nabla^2 f(x_j)$

→ uses $\nabla^2 f(x_j)$ in line search to get α_j.

→ Numerical issues w/ small #s.

Rank two Correction – DFP

→ update $\nabla^2 f(x_j)$ w/ rank 2 matrix maintaining symmetry, +ve def and (*).

→ $\{(g_{k+1} - g_k)\nabla^2 f(x_k)^{-1}\}$ produced are $\nabla^2 f$ conjugate; is a conj dir method!

↳ Convergence – n-steps [Quadratic]

Broyden Family

– All produce same x_{k+1} if $\alpha_k = \min_{\alpha}(f_k + \alpha d_k)$

→ satisfy $(g(x_{k+1}) - g(x_k)) = \nabla^2 f(x_k)(x_{k+1} - x_k)$ instead of (*)

→ update approx with complimentary formula (swap $g(x_j)$ w/ x_j's), or take inverse of $\nabla^2 f(x_k)^{-1}$.

→ Both BFGS and DFP have symmetric rank 2 corrections, so $\nabla^2 f(x)^{\phi}$ symmetric, and satisfies

$\nabla^2 f(x)^{\phi} = (1-\phi)\nabla^2 f(x)^{DFP} + \phi \nabla^2 f(x)^{BFGS}$

↳ preserves symmetry, (*), if $\phi \ge 0$ also preserves +ve def.

Pure – uses fixed ϕ.

– Convergence – superlinear when viewed in blocks of n steps, which approx 1 step of conj. dir. ⇒ Convergence may be slow if n large.

– Sensitive to line search inaccuracies.

Impure – uses sequence of $\{\phi_k\}$

General Line Search Algorithm $S: E^{2n} \to P(E^n)$, $S(\vec{x},\vec{d}) = \{\vec{y} \mid \vec{y} = \vec{x} + \alpha\vec{d},\ \alpha \geq 0,\ f(y) = \min_{0 \leq \alpha < \infty} f(\vec{x}+\alpha\vec{d})\}$

→ Minimize over a line in direction of travel $\vec{d}$. Assume unimodal.

Curve fitting → minimize over polynomial approximation of curve.
→ locally convergent

Newton Method → $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$
→ $f \in C^2$, $f''(x_k) \neq 0$, $f(x_k) = 0$ or else won't converge.

Method of false position → replace $f''(x_k)$ with $\frac{f(x_{k-1}) - f(x_k)}{x_{k-1} - x_k}$.
→ Same convergence as N.M
→ $f \in C^1$

Cubic fit → Minimize over cubic approximation of curve.
→ faster convergence than above but needs $f'(x_k)$, ($f \in C^1$).

Quadratic fit → Minimize over quadratic approximation of curve.
→ does not use $f'(x_k)$, only x_{k-2}, x_{k-1}, x_k
→ Can be globally convergent if use 3-pt pattern: (x_1, x_2, x_3) where $x_1 < x_2 < x_3 \wedge f(x_1), f(x_3) \geq f(x_2)$
Use QF and maintain pattern to ensure convergence.

Stopping criteria → percentage test, Armijo test, etc...
→ determine what is too high and too low for estimate of α.