Optimization (non-linear)

Hugo SALOU

Steepest descent

$\vec{p}_k = -\underbrace{\nabla f(\vec{x}_k)}_{\vec{g}_k}$

$\alpha_k = \vec{g}_k^T \vec{g}_k / \vec{g}_k^T A \vec{g}_k$ (for quadratic functions)

Newton's method

$H_f(\vec{x}_k)\,\vec{p}_k = -\nabla f(\vec{x}_k)$

Rayleigh Quotient

$r_A(\vec{v}) = \vec{v}^T A \vec{v} / \vec{v}^T\vec{v}$ (symmetric A)

$\lambda_{min}(A) \leq r_A(\vec{v}) \leq \lambda_{max}(A)$

Quasi Newton Method

BFGS

While $\|\nabla f(\vec{x}_k)\| > \varepsilon$ do

- $\vec{p}_k \leftarrow -\tilde{B}_k \nabla f(\vec{x}_k)$, with $\tilde{B}_k$ approx of $(H_f(\vec{x}_k))^{-1}$
- step w/ line search
- $\tilde{B}_{k+1} \leftarrow (I - \rho_k \vec{s}_k \vec{y}_k^T)\tilde{B}_k \times (I - \rho_k \vec{y}_k \vec{s}_k^T) + \rho_k \vec{s}_k \vec{s}_k^T$

where
$\vec{s}_k = \vec{x}_{k+1} - \vec{x}_k$
$\vec{y}_k = \nabla f(\vec{x}_{k+1}) - \nabla f(\vec{x}_k)$
$\rho_k = 1/\vec{y}_k^T \vec{s}_k > 0$

This converges assuming
- $L := \{\vec{x} \mid f(\vec{x}) \leq f(\vec{x}_0)\}$ is convex
- $\exists m, M$, st $m\|\vec{y}\|^2 \leq \vec{y}^T H_f(\vec{x})\vec{y} \leq M\|\vec{y}\|^2$ $\forall \vec{y} \in \mathbb{R}^n, \forall x \in L.$

Least squares

$$\min_{\vec{x}} \underbrace{\tfrac{1}{2}\|\vec{F}(\vec{x})\|^2}_{f(\vec{x})}$$

It holds

$\nabla f(\vec{x}) = J(\vec{x})^T \vec{F}(\vec{x})$

$H_f(\vec{x}) = J(\vec{x})^T J(\vec{x}) + \sum_i F_i(\vec{x}) H_{F_i}(\vec{x})$

where
$$J(\vec{x}) = \begin{pmatrix} \nabla F_1(\vec{x}) \\ \vdots \\ \nabla F_m(\vec{x}) \end{pmatrix}$$

if $\vec{F}(\vec{x}) = A\vec{x} + \vec{b}$ with A constant then $\nabla f(\vec{x}) = A^T(A\vec{x} + \vec{b})$ and $H_f(\vec{x}) = A^T A$ and thus $\vec{x}^*$ global min iff $A^T A \vec{x} = -A^T \vec{b}$.

normal equations.

Armijo & Wolf conditions

Armijo's rule is
$$f(\vec{x}_k + \alpha_k \vec{p}_k) \leq f(\vec{x}_k) + \alpha_k c_1 \nabla f(\vec{x}_k)\,\vec{p}_k \quad (A)$$
where $c_1 \in (0,1)$.

⤳ avoid steps too large

Wolfe's rule is
$$\nabla f(\vec{x}_k + \alpha_k \vec{p}_k)^T \vec{p}_k \geq c_2 \nabla f(\vec{x}_k)^T \vec{p}_k \quad (W)$$
where $c_2 \in (c_1, 1)$

⤳ avoids steps too large

Convexity & minima

Stationary point: $\nabla f(\vec{x}) = \vec{0}$

$\vec{x}$ local min of f iff $\nabla f(\vec{x}) = \vec{0}$ and $H_f(\vec{x}) \succcurlyeq 0$

$\Downarrow$ f convex

$\vec{x}$ global min of f

strictly convex $\Longrightarrow$ nb. min ≤ 1

Eigenvalues $(\lambda_i)_i$ of $H_f(\vec{x})$ with $\nabla f(\vec{x}) = \vec{0}$.

$\forall i, \lambda_i > 0 \longrightarrow$ local min
$\forall i, \lambda_i < 0 \longrightarrow$ local max
$\forall i, \lambda_i = 0 \longrightarrow$ degenerate pt.
$\exists i, j, \lambda_i > 0 \;\&\; \lambda_j < 0 \longrightarrow$ saddle pt.

Convex: $f(\vec{x} + t(\vec{y} - \vec{x})) \leq f(\vec{x}) + t(f(\vec{y}) - f(\vec{x}))$
i.e. $f(\vec{y}) \geq f(\vec{x}) - \nabla f(\vec{x})^T(\vec{x} - \vec{y})$

$\Uparrow$

strictly convex: $f(\vec{x} + t(\vec{y} - \vec{x})) < f(\vec{x}) + t(f(\vec{y}) - f(\vec{x}))$
i.e. $f(\vec{y}) > f(\vec{x}) - \nabla f(\vec{x})^T(\vec{x} - \vec{y})$

$\Uparrow$

μ-strongly convex: $f(\vec{y}) \geq f(\vec{x}) - \nabla f(\vec{x})^T(\vec{x} - \vec{y}) + \frac{\mu}{2}\|\vec{x} - \vec{y}\|^2$
i.e. $\vec{x} \mapsto f(\vec{x}) - \frac{\mu}{2}\|\vec{x}\|^2$ is convex

Gauss-Newton method

Quasi Newton w/ $B_k = J(\vec{x}_k)^T J(\vec{x}_k)$.

$J(\vec{x}_k)^T J(\vec{x}_k)\,\vec{p}_k = -J(\vec{x}_k)^T \vec{F}(\vec{x}_k)$

assumes $J(\vec{x}_k)$ full rank.

Levenberg-Marquardt method

$$(J(\vec{x}_k)^T J(\vec{x}_k) + \lambda_k I)\,\vec{p}_k = -J(\vec{x}_k)^T F(\vec{x}_k)$$

λ_k: regularization parameter

Quadratic functions

$q(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T\vec{x}$ (A symmetric)

$\nabla q(\vec{x}) = A\vec{x} - \vec{b}$ and $H_q(\vec{x}) = A$.

Line Search

(1) choose $\vec{p}_k$
(2) find α_k st (A) & (W)
(3) iterate until $\|\nabla f(\vec{x}_{k+1})\| \leq \varepsilon$

fix α_0, $c_1 \in (0,1)$ and $\gamma \in (0,1)$
if $f(\vec{x}_k + \alpha_k \vec{p}_k) \leq f(\vec{x}_k) + \alpha_k c_1 \nabla f(\vec{x}_k)^T \vec{p}_k$ then Stop!
otherwise $\alpha_k \leftarrow \alpha_k \times \gamma$ and repeat (for a fixed number of iterations)

Lagrangian

Search Space $K = \left\{ \vec{x} \;\middle|\; \begin{array}{l} \forall i, h_i(\vec{x}) = 0 \\ \forall j, g_j(\vec{x}) \leq 0 \end{array} \right\}$
where $(\nabla h_i(\vec{x}))$ are lin. indep

$$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu}) = f(\vec{x}) + \sum_i \lambda_i h_i(\vec{x}) + \sum_j \mu_j g_j(\vec{x})$$

KKT

When f and g_j convex
$\vec{x}^*$ is a global min of f on K
iff
$\exists \vec{\lambda}^*, \vec{\mu}^*$,
- $\nabla \mathcal{L}(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) = \vec{0}$
- $h_i(\vec{x}^*) = \vec{0}$
- $g_j(\vec{x}^*) \geq 0$
- $\mu_j^* \geq 0$
- $\mu_j^* g_j(\vec{x}^*) = 0$ — "activity" of constraints

descent direction

$\vec{p}$ for f in $\vec{x}$ if
$$\frac{\partial f}{\partial \vec{p}}(\vec{x}) = \nabla f(\vec{x})^T \vec{p} < 0.$$

Taylor formula

$$f(\vec{x} + \vec{h}) = f(\vec{x}) + \nabla f(\vec{x} + t\vec{h})^T \vec{h}$$
for some $t \in (0,1)$

$$f(\vec{x} + \vec{h}) = f(\vec{x}) + \nabla f(\vec{x})^T \vec{h} + \tfrac{1}{2}\vec{h}^T H_f(\vec{x} + t\vec{h})\vec{h}$$

Optimization (linear)

Hugo BALOU

Weak duality

any val in (D)
$\geq$
any val in (P)

If (P) has an optimal sol°, then so does (D)

and opt(P) = opt(D)

Strong duality

$Ax \leq b$ has no sol°

iff

a non-negative combination of these inequations give $0 \leq -1$

Complementary Slackness

given a potential optimal sol° x

- if $\mathrm{Cond}_i^{(P)}(x)$ is strict then $y_i = 0$
- conversely, if $x_j \neq 0$, $\mathrm{Cond}_j^{(D)}$ is an equality

Then we solve the system.

Separation Oracle for P polyhedron

given $x \in \mathbb{R}^n$, returns either

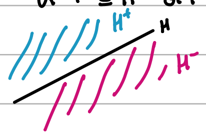
- True when $x \in P$
- Otherwise valid constraints st. $a^T x \leq b$ & $a^T x > b$

Definitions

- polytope: Convex Hull (CH) of finite $\subseteq \mathbb{R}^n$
- polyhedron: finite intersection of half spaces

bounded polyhedra = polytopes

- $\dim(P) = \max_{CH \subseteq P} (\dim CH)$
  $= \min_{P \subseteq A} (\dim A)$
  A affine space
- face of P: $H \cap P \subseteq P$
  when H is a hyperplane
  & $P \subseteq H^+$ or $P \subseteq H^-$



Rounding

good for approximation

deterministic vs randomized rounding

Totally Unimodular Matrices

- Seymour: checking is in P
- Examples
  - inc. matrix of bipartite graphs (0/1)
  - inc. matrix of oriented graphs (-1,0,1)
  - 0,1 matrix where 1's are consecutive in columns.
  - ⚠ network matrices

Simplex algorithm

Phase I: illegal pivot & pivots

Phase II: more pivots

If cycle, give up!

$\boxed{3 \times 3 = 9}$