Starting 11/16

General Updates

- Next Semester: Nonlinear II [553.762]

  (Constrained Optimality Conditions, Duality, Linear/Quad/Semidefinite Programming, Interior Point Methods)

- Next Semester: Nonsmooth OPT Seminar [553.861]

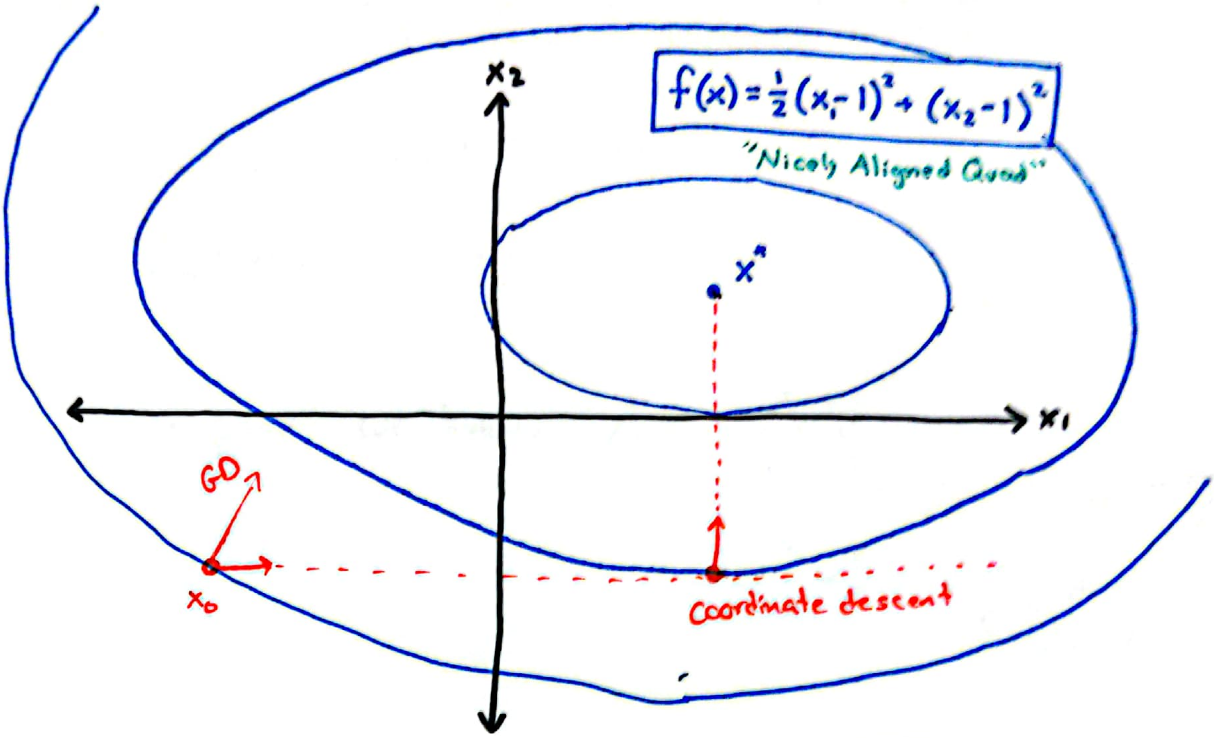
  (Reading Course for [AMS] PhD students on nonsmooth, nonconvex calculus following Clarke's book "Nonsmooth Analysis and Control Theory")

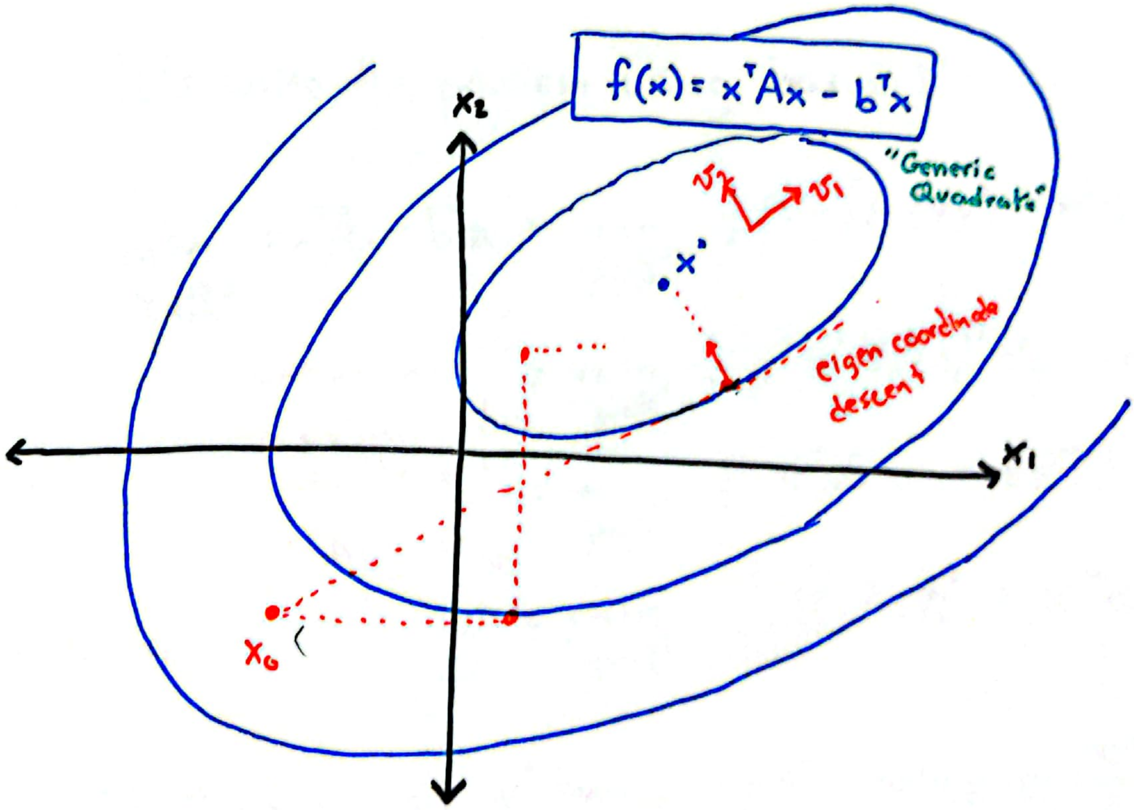
- <u>Please fill out course evaluation</u>

Today: Conjugate Gradient Methods for Least Squares

1. Easy if we know Spectral Decomposition
2. Conjugate Vectors
3. Conjugate Gradient Method
4. Convergence Guarantees

Two motivating quadratics





1. Easy case for $\min f(x) = \frac{1}{2}x^T A x - b^T x$, $A \succ 0$

Captures Least Squares

Captures solving $B_k p = -\nabla f(x_k)$

Suppose we know $A = V \Lambda V^T$

$V = (v_1\ v_2 \cdots v_d)$

$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$

Lets write the problem using basis V:

$$\min_{x \in \mathbb{R}^d} \frac{1}{2}x^T A x - b^T x = \min_{y \in \mathbb{R}^d} \frac{1}{2}(Vy)^T A (Vy) - b^T (Vy)$$

$$= \min_{y \in \mathbb{R}^d} \frac{1}{2} y^T V^T V \Lambda V^T V y - b^T (Vy)$$

$$= \min_{y \in \mathbb{R}^d} \frac{1}{2} y^T \Lambda y - (V^T b)^T y$$

$$= \min_{y \in \mathbb{R}^d} \sum_{i=1}^{d} \left(\frac{1}{2} \lambda_i y_i^2 - v_i^T b\, y_i \right)$$

Working in V the problem seperates since $v_i^T A v_j = 0$ $\forall i \neq j$

conjugacy

$\Rightarrow$ Minimized at $y_i^* = \operatorname{argmin} \frac{\lambda_i}{2} y_i^2 - v_i^T b\, y_i = \frac{v_i^T b}{\lambda_i}$.

$\Rightarrow$ Minimized f at $x^* = \sum v_i y_i^* = \sum_{i=1}^{d} \frac{v_i v_i^T b}{\lambda_i}$.

2. Conjugate Vectors

Lets define a second inner product based on $A \succ 0$

$$\langle x, y \rangle_A = x^T A y$$

$\langle x,x \rangle \geq 0 \ \forall x$, $\langle x,x \rangle = 0$ iff $x=0$
$\langle x,y \rangle = \langle y,x \rangle$
$\langle x+y, z \rangle = \langle x,z \rangle + \langle y,z \rangle$
$\langle \alpha x, y \rangle = \alpha \langle x,y \rangle$

Our normal product is

$$\langle x, y \rangle_I = x^T y.$$

Definition 1. x, y are A-conjugate if $\langle x,y \rangle_A = 0$
(that is, they are orthogonal w.r.t. A)
"with respect to"

2. Given a linear subspace $L \subseteq \mathbb{R}^d$, the orthogonal complement w.r.t. A is

$$L_A^{\perp} = \{ y \mid \langle x,y \rangle_A = 0 \ \forall x \in L \}$$

3. The projection of x onto y w.r.t. A is

$$\frac{\langle x,y \rangle_A}{\langle y,y \rangle_A} y.$$

Lemma Let $x^1 \dots x^k$ that are A-conjugate, then they linearly independent and $x \in \text{span}(x^1 \dots x^k)$ has unique decomposition

$$x = \frac{\langle x, x^1 \rangle_A}{\langle x^1, x^1 \rangle_A} x_1 + \dots + \frac{\langle x, x^k \rangle_A}{\langle x^k, x^k \rangle_A} x^k.$$

We want a set of A-conjugate vector. spanning $\mathbb{R}^d$.

Eigenvectors work but are expensive to get.

Classic Construction : Gram-Schmidt Orthogonalization

(QR decomposition)

Input: $A \succ 0$, and linearly independent $x^1 \dots x^k$

Output: vectors $s^1 \dots s^k$ that are A-conjugate

with $\text{span}(s^1 \dots s^k) = \text{span}(x^1 \dots x^k)$

$$s^1 = x^1$$

$$s^{i+1} = x^{i+1} - \sum_{j=1}^{i} \frac{\langle x^{i+1}, s^j \rangle_A}{\langle s^j, s^j \rangle_A} s^j$$

Check $\text{span}(s^1 \dots s^k) = \text{span}(x^1 \dots x^k)$. Inductively true.

Check $\langle s^{i+1}, s^j \rangle_A = 0 = \langle s^j, x^{i+1} \rangle_A - \sum_{n=1}^{i} \frac{\langle x^{i+1}, s^n \rangle_A}{\langle s^n, s^n \rangle_A} \langle s^n, s^j \rangle_A$

$$= \langle s^j, x^{i+1} \rangle_A - \frac{\langle x^{i+1}, s^j \rangle_A}{\langle s^j, s^j \rangle_A} \langle s^j, s^j \rangle_A$$

$$= 0.$$

Check $s^{i+1} \neq 0$. Linear Independence of x^k rules this out.

So if you provide linearly ind vectors $x^1 \dots x^d$, then

I can construct A-conjugate vectors $s^1 \dots s^d$

(also linearly independent).

$$S = (s_1 \cdots s_d)$$

$$\min_x \tfrac{1}{2} x^T A x - b^T x = \min_y \tfrac{1}{2} (Sy)^T A (Sy) - b^T (Sy)$$

$$= \min_y \sum_{ij} \tfrac{1}{2} y_i \underline{s_i^T A s_j} y_j - b^T s_i y_i$$

$$= \min_y \sum_i \left(\tfrac{1}{2} y_i^2 s_i^T A s_i - b^T s_i y_i \right)$$

$$\Rightarrow y_i^* = \frac{b^T s_i}{s_i^T A s_i} \quad \text{minimizes}$$

$\Rightarrow$ Original problem minimizes at

$$x^* = \sum_{i=1}^{d} \frac{s_i s_i^T b}{s_i^T A s_i}.$$

Computation still $O(d^3)$. $O(d)$ Gram-Schmidt steps using $O(d^2)$ for matrix vector multiply $\langle x, y \rangle_A$.

3. Conjugate Gradient Method

Lets use gradients to build a smart basis to work from. (Hope to get near optimal in few steps. stop early).

<u>Lemma 1</u> For any x^0 and $s^1 \ldots s^k$, consider the subspace restricted minimization

$$\min \ \tfrac{1}{2}x^TAx - b^Tx = f(x)$$
$$\text{s.t. } x \in x^0 + \operatorname{span}(s^1 \ldots s^k).$$

Then the minimizer $\bar{x}$ has $\nabla f(\bar{x})$ orthogonal to $\operatorname{span}(s^1 \ldots s^k)$.

(using our old inner product)

Proof. Equivalent to unconstrained min

$$\min_y \ f(x^0 + Sy)$$

The optimality condition says $S^T \nabla f(x^0 + Sy^*) = 0$.

This says $\nabla f(\bar{x})$ is orthogonal to each s_i. $\square$

Solving restricted problems gives a gradient that is not dependent on previous $s^1 \ldots s^k$.

Lemma 2 Suppose $\bar{x}$ minimizes f over $x^0 + \mathrm{span}(s^1 \dots s^k)$ and s^{k+1} is A-conjugate to each s^i. Then

$\hat{x} = \underset{x = \bar{x} + \alpha s^{k+1}}{\mathrm{argmin}} \; f(x)$ is the minimizer of f over $\mathrm{span}(s^1, \dots, s^{k+1}) + x_0$.

Proof. Essential same as previous seperation we have seen. $\square$

This gives the Conjugate Gradient Method

Given $x_0 \in \mathbb{R}^d$, $s^0 = r^0 = b - Ax^0 = -\nabla f(x^0)$.

Iterate:

$$\alpha_i = \underset{\alpha}{\mathrm{argmin}} \; f(x^i + \alpha s^i) \leftarrow \alpha_i = \frac{s^{iT}(b - Ax^i)}{\langle s^i, s^i \rangle_A}.$$

$$x^{i+1} = x^i + \alpha_i s^i$$

$$r^{i+1} = -\nabla f(x^{i+1}) = b - Ax^{i+1}$$

$$s^{i+1} = r^{i+1} - \sum_{j=1}^{i} \frac{\langle r^{i+1}, s^j \rangle_A}{\langle s^j, s^j \rangle_A} s^j.$$