Algorithms for large scale structured optimization problems

First-order methods for optimization, variational-inequality and saddle-point problems

(Second Lecture)

Renato D.C. Monteiro
(Georgia Tech)

Fields 09 – Toronto
October 20-24, 2009

Outline for the second lecture

- **Smooth convex optimization**
  - Functions with Lips. continuous gradient
  - Projected gradient method
  - Lower complexity bound
  - Nesterov's optimal method and its variants
  - Application to cone programming
  - Extensions (smooth + nonsmooth funct's)
- **Nonsmooth convex optimization**
  - Lower complexity bound
  - Subgradient method
- **Specially structured convex optimization**
  - Problem class
  - Nesterov's smooth approximation scheme
  - Remarks
- **Algorithms for V.I. and Saddle Pt. Probl.'s**
  - V.I. and the saddle pt. problem
  - Prox-mirror and Korpelevich's methods
  - Complexity results for V.I.
  - Complexity results for S.P. problems

Smooth convex optimization

Definition: Let $\mathbf{X} \subseteq \Re^n$ be a convex set. A differentiable convex function $\mathbf{f} : \mathbf{X} \to \Re$ has $\mathbf{L}$-Lipschitz continuous gradient if

$$\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\|_* \leq \mathbf{L}\|\mathbf{y} - \mathbf{x}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}.$$

where $\|\phi\|_* = \max\{\phi(\mathbf{x}) : \|\mathbf{x}\| \leq \mathbf{1}\}$. We write $\mathbf{f} \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}}(\mathbf{X})$.

Notation: For $\mathbf{x} \in \mathbf{X}$, let $\mathbf{l_f}(\cdot; \mathbf{x})$ denote the first-order approximation of $\mathbf{f}$ at $\mathbf{x}$:

$$\mathbf{l_f}(\mathbf{y}; \mathbf{x}) := \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in \Re^n$$

Proposition: The following are equivalent:

a) $\mathbf{f} \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}}(\mathbf{X})$;

b) $\mathbf{f} \leq \mathbf{l_f}(\cdot; \mathbf{x}) + \mathbf{L}\| \cdot - \mathbf{x}\|^2/2$;

Projected gradient method

Given "simple" convex $X \subseteq \Re^n$ **and** $\mathbf{f} \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}}(\mathbf{X})$**, consider the problem**

$$\mathbf{f}^* := \min\{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$$

and let $\mathbf{X}^* \neq \emptyset$ **denote its set of optimal sol's. Assume** $\|\cdot\|$ **denote an inner product norm.**

Projected gradient method (with fixed stepsize)

0) **Let** $\mathbf{x_0} \in \mathbf{X}$ **and** $\alpha > \mathbf{0}$ **be given. Set** $\mathbf{k} = \mathbf{0}$.

1) **Compute**

$$\mathbf{x_{k+1}} := \mathrm{argmin}\left\{\alpha\, \mathbf{l_f}(\mathbf{x}; \mathbf{x_k}) + \frac{\mathbf{L}}{\mathbf{2}}\|\mathbf{x} - \mathbf{x_k}\|^{\mathbf{2}} : \mathbf{x} \in \mathbf{X}\right\} \quad (*)$$

2) **Set** $\mathbf{k} \leftarrow \mathbf{k} + \mathbf{1}$ **and go to step 1.**

Iteration $(*)$ **can also be written as**

$$\mathbf{x_{k+1}} := \mathbf{P_X}\left(\mathbf{x_k} - \frac{\alpha}{\mathbf{L}}\mathbf{f}'(\mathbf{x_k})\right)$$

where $\mathbf{P_X}$ **denotes the projection operator onto** $\mathbf{X}$**.**

Proposition: **If** $\alpha \in (\mathbf{0}, \mathbf{2})$**, then**

$$\mathbf{f}(\mathbf{x_k}) - \mathbf{f}^* \leq \frac{\mathbf{L}\|\mathbf{x_0} - \mathbf{x}^*\|^{\mathbf{2}}}{\alpha(\mathbf{2} - \alpha)\,\mathbf{k}}$$

for any $\mathbf{x}^* \in \mathbf{X}^*$**.**

Lower-complexity bound

Theorem: For any $\mathbf{1 \leq k \leq (n-1)/2}$ and $\mathbf{x_0 \in \Re^n}$, there exists quadratic function $\mathbf{f} \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}}(\Re^{\mathbf{n}})$ with the following property: any first-order method such that

$$\mathbf{x_k \in x_0 + \mathrm{lin}\{f'(x_0), \ldots, f'(x_{k-1})\}}$$

for solving $\min\{\mathbf{f(x) : x \in \Re^n}\}$, satisfies

$$\mathbf{f(x_k) - f^* \geq \frac{3L\|x_0 - x^*\|^2}{32(k+1)^2}}$$

Bregman's distance: Let $\mathbf{h : X \to \Re}$ be a differentiable σ-strongly convex function, i.e.:

$$\mathbf{h(\cdot) \geq l_h(\cdot; x) + \frac{\sigma}{2}\|\cdot - x\|^2, \quad \forall x \in X}$$

Define the Bregman distance $\mathbf{d_h : X \times X \to \Re}$ as

$$\mathbf{d_h(\tilde{x}; x) = h(\tilde{x}) - l_h(\tilde{x}; x), \quad \forall x, \tilde{x} \in X}$$

Obs: If $\mathbf{h = \|\cdot\|^2/2}$, then $\mathbf{d_h(\tilde{x}; x) = \|\tilde{x} - x\|^2/2}$.

Nesterov's optimal method

Let $\{\alpha_k\} \subseteq \Re_{++}$ be such that

$$\alpha_0 \in (0,1], \qquad \alpha_k^2 \leq \sum_{i=0}^{k} \alpha_i$$

and set $\tau_k = \alpha_k / \sum_{i=0}^{k} \alpha_i$. For example, $\alpha_k = (k+1)/2$ and $\tau_k = 2/(k+2)$.

Nesterov's algorithm:

0) Let $x_0 \in X$ be given and set $u_0 = x_0$ and $k = 1$.

1) Using $u_0, \ldots, u_{k-1}$, compute

$$\begin{aligned} u_k^a &:= \arg\min_{x \in X} \frac{1}{L} \sum_{i=0}^{k-1} \alpha_i \, l_f(x; u_i) + \frac{1}{\sigma} d_h(x; u_0) \\ x_k &:= \arg\min_{x \in X} \frac{1}{L} l_f(x; u_{k-1}) + \frac{1}{2} \|x - u_{k-1}\|^2 \qquad (*) \\ u_k &:= \tau_k u_k^a + (1 - \tau_k) x_k \end{aligned}$$

2) Set $k \leftarrow k+1$ and go to step 1.

Tseng's variant: Replace $(*)$ by

$$x_k = (1 - \tau_{k-1}) x_{k-1} + \tau_{k-1} u_k^a$$

Nesterov's optimal method (cont.)

Proposition: For every $\mathbf{k}$ and $\mathbf{x^* \in X^*}$:

$$\mathbf{f(x_k) - f^* \leq \frac{L\, d_h(x^*; x_0)}{\sigma \sum_{i=0}^{k-1} \alpha_i}}$$

In particular, if $\alpha_\mathbf{k} = (\mathbf{k}+\mathbf{1})/\mathbf{2}$ for all $\mathbf{k}$, then

$$\mathbf{f(x_k) - f^* \leq \frac{4L\, d_h(x^*; x_0)}{\sigma k(k+1)} = \mathcal{O}\left(\frac{1}{k^2}\right)}$$

Lower bound: Assume $\mathbf{X}$ is *bounded* and let $\theta_\mathbf{k}$ and $\mathbf{v_k}$ be the optimal value and optimal solution of

$$\frac{\mathbf{1}}{\sum_{\mathbf{i=0}}^{\mathbf{k-1}} \alpha_\mathbf{i}} \min \left\{ \sum_{\mathbf{i=0}}^{\mathbf{k-1}} \alpha_\mathbf{i}\, \mathbf{l_f(x; u_i) : x \in X} \right\}$$

Proposition: For any $\mathbf{k \geq 0}$,

$$\mathbf{f(x_k) - f^* \leq f(x_k) - \theta_k \leq \frac{L\, d_h(v_k; x_0)}{\sigma \sum_{i=0}^{k-1} \alpha_i}}$$

Observation: In practice, the number of iterations is usually proportional to the theoretical bound. Hence, if $\mathbf{L}$ is too large and/or $\mathbf{x_0}$ is far from $\mathbf{x^*}$, convergence can be quite slow.

Auslender-Teboule's variant: Let $\alpha_k = (k+2)/2$ and $\tau_k = 1/\alpha_k$ for all k.

0) Given $x_0 \in X$, set $u_0^a = x_0$ and $k = 1$.

1) Using $u_{k-1}^a, x_{k-1}, u_{k-1}$, compute

$$
\begin{aligned}
u_k^a &:= \arg\min_{x \in X} \frac{1}{L}\alpha_{k-1}\, l_f(x; u_{k-1}) + \frac{1}{\sigma} d_h(x; u_{k-1}^a) \\
x_k &:= (1 - \tau_{k-1})x_{k-1} + \tau_{k-1} u_k^a \\
u_k &:= \tau_k u_k^a + (1 - \tau_k) x_k
\end{aligned}
$$

2) Set $k \leftarrow k + 1$ and go to step 1.

Proposition: For every k and $x^* \in X^*$:

$$
f(x_k) - f^* \le \frac{4L d_h(x^*; x_0)}{\sigma k(k+2)} = \mathcal{O}\left(\frac{1}{k^2}\right)
$$

Applications to Cone Progr.

Consider

$$\min\{\langle \mathbf{c}, \mathbf{x}\rangle : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \in \mathcal{K}\},$$
$$\max\{\langle \mathbf{b}, \mathbf{y}\rangle : \mathbf{A}^*\mathbf{y} + \mathbf{s} = \mathbf{c},\ \mathbf{s} \in \mathcal{K}^*\}$$

where $\mathcal{K}$ is a closed convex cone and

$$\mathcal{K}^* = \{\mathbf{s} : \langle\, \mathbf{s}, \mathbf{x}\,\rangle \geq \mathbf{0},\ \forall \mathbf{x} \in \mathcal{K}\}$$

Assuming that there is no duality, we can solve the above pair of dual problems by means of the following smooth reformulation

$$\begin{array}{ll} \min & (\langle \mathbf{c}, \mathbf{x}\rangle - \langle \mathbf{b}, \mathbf{y}\rangle)^{\mathbf{2}} + \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^{\mathbf{2}} + \|\mathbf{A}^*\mathbf{y} + \mathbf{s} - \mathbf{c}\|^{\mathbf{2}} \\ \textbf{s.t.} & (\mathbf{x}, \mathbf{s}) \in \mathcal{K} \times \mathcal{K}^* \end{array}$$

or alternatively,

$$\begin{array}{ll} \min & [\mathbf{d}_{\mathcal{K}}(\mathbf{x})]^{\mathbf{2}} + [\mathbf{d}_{\mathcal{K}^*}(\mathbf{s})]^{\mathbf{2}} \\ \textbf{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A}^*\mathbf{y} + \mathbf{s} = \mathbf{c}, \quad \langle \mathbf{c}, \mathbf{x}\rangle = \langle \mathbf{b}, \mathbf{y}\rangle \quad (**) \end{array}$$

See Lan, Lu, M. (2009), Jarre and Rendl (2008).

Smooth+Nonsmooth Functions

Assume that $\mathbf{f(x) = f_S(x) + f_N(x)}$, where $\mathbf{f_S} \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}}(\mathbf{X})$ and $\mathbf{f_N : X \to \Re}$ is a closed convex function.

There exist extensions of Nesterov's algorithm and its variants, which instead of using the linear approximation $\mathbf{f(x) \approx l_f(x; u_k)}$, use

$$\mathbf{f(x) \approx l_{f_s}(x; u_k) + f_N(x)}$$

This leads to subproblems of the form

$$\min_{\mathbf{x \in X}} \langle \mathbf{c, x} \rangle + \mathbf{f_N(x)} + \tau \mathbf{d_h(x; u_0)}$$

for some $\mathbf{c} \in \Re^{\mathbf{n}}$ and $\tau > \mathbf{0}$.

Exactly the same complexity bounds can be derived for these extensions.

Example: **If $\mathbf{f(x)} = \frac{\mathbf{1}}{\mathbf{2}}\|\mathbf{Ax - b}\|^{\mathbf{2}} + \tau\|\mathbf{x}\|_{\mathbf{1}}$ and $\mathbf{h} = \|\cdot\|^{\mathbf{2}}/\mathbf{2}$, then the above subproblem has a closed form solution (see Wright et al.)**

Algorithms for Nonsmooth Functions

Consider the problem

$$\mathbf{f}^* := \min\{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$$

where $\mathbf{X} \subseteq \Re^{\mathbf{n}}$ is closed convex and $\mathbf{f} : \Re^{\mathbf{n}} \to \Re$ is convex. Let $\mathbf{X}^* \neq \emptyset$ denote its set of optimal sol's.

Lower complexity bound: Assume $\mathbf{X} = \Re^{\mathbf{n}}$. Given $\mathbf{x_0} \in \Re^{\mathbf{n}}$ and positive constants $\mathbf{R}, \mathbf{M}$, consider the class $\mathcal{F}(\mathbf{x_0}, \mathbf{M}, \mathbf{R})$ of functions $\mathbf{f}$ such that:

a) $\exists\, \mathbf{x}^* \in \mathbf{X}^*$ such that $\|\mathbf{x}^* - \mathbf{x_0}\| \leq \mathbf{R}$;

b) $\mathbf{f}$ is $\mathbf{M}$-Lipschitz continuous on the closed ball $\{\mathbf{x} : \|\mathbf{x} - \mathbf{x_0}\| \leq \mathbf{R}\}$.

Proposition: For any $\mathbf{k} \leq \mathbf{n} - \mathbf{1}$, there exists $\mathbf{f} \in \mathcal{F}(\mathbf{x_0}, \mathbf{M}, \mathbf{R})$ with the property that any algorithm which generates $\{\mathbf{x_k}\}$ such that

$$\mathbf{x_k} \in \mathbf{x_0} + \mathrm{lin}\{\mathbf{g_0}, \ldots, \mathbf{g_{k-1}}\},$$

where $\mathbf{g_i} \in \partial \mathbf{f}(\mathbf{x_i})$ for all $\mathbf{i}$, satisfies

$$\mathbf{f}(\mathbf{x_k}) - \mathbf{f}^* \geq \frac{\mathbf{MR}}{\mathbf{2}(\mathbf{1} + \sqrt{\mathbf{k} + \mathbf{1}}\,)}$$

Question: Is there an optimal method?

Subgradient method

Subgradient method:

0) Let $\mathbf{x_0} \in \mathbf{X}$ be given and set $\mathbf{k} = \mathbf{0}$.

1) Choose $\alpha_\mathbf{k} > \mathbf{0}$ and $\mathbf{g_k} \in \partial \mathbf{f}(\mathbf{x_k})$, and set

$$\mathbf{x_{k+1}} := \mathbf{P_X}\left(\mathbf{x_k} - \alpha_\mathbf{k} \frac{\mathbf{g_k}}{\|\mathbf{g_k}\|}\right)$$

2) Set $\mathbf{k} \leftarrow \mathbf{k} + \mathbf{1}$ and go to step 1.

Proposition: Assume that $\mathbf{f}$ is $\mathbf{M}$-Lipschitz continuous on $\mathbf{X}$. Then, for all $\mathbf{k}$ and $\mathbf{x}^* \in \mathbf{X}^*$

$$\min_{\mathbf{i}=\mathbf{0},\ldots,\mathbf{k}} [\mathbf{f}(\mathbf{x_i}) - \mathbf{f}^*] \leq \mathbf{M}\left(\frac{\|\mathbf{x_0} - \mathbf{x}^*\|^2 + \sum_{\mathbf{i}=\mathbf{0}}^{\mathbf{k}} \alpha_\mathbf{i}^2}{2\sum_{\mathbf{i}=\mathbf{0}}^{\mathbf{k}} \alpha_\mathbf{i}}\right)$$

Corollary: Fix $\mathbf{K} \geq \mathbf{0}$. If $\mathbf{R} \geq \|\mathbf{x_0} - \mathbf{x}^*\|$ is known and we set $\alpha_\mathbf{k} = \mathbf{R}/\sqrt{\mathbf{K}+\mathbf{1}}$ for all $\mathbf{k} = \mathbf{0}, \ldots, \mathbf{K}$, then

$$\min_{\mathbf{i}=\mathbf{0},\ldots,\mathbf{K}} [\mathbf{f}(\mathbf{x_i}) - \mathbf{f}^*] \leq \frac{\mathbf{MR}}{\sqrt{\mathbf{K}+\mathbf{1}}}$$

Specially Structured Convex Optim.

Consider $\mathbf{f}^* := \min\{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$, **where**

$$\mathbf{f}(\mathbf{x}) = \phi(\mathbf{x}) + \max_{\tilde{\mathbf{x}} \in \tilde{\mathbf{X}}} \langle \tilde{\mathbf{x}}, \mathbf{C}\mathbf{x} \rangle - \tilde{\phi}(\tilde{\mathbf{x}}),$$

$\phi \in \mathrm{Conv}^{\mathbf{1}}_{\mathbf{L}_\phi}(\mathbf{X})$, **set** $\tilde{\mathbf{X}} \subseteq \Re^{\tilde{\mathbf{n}}}$ **is compact convex,** $\mathbf{C} : \Re^{\mathbf{n}} \to \Re^{\tilde{\mathbf{n}}}$ **is linear, and** $\tilde{\phi} : \tilde{\mathbf{X}} \to \Re$ **is convex**

The dual problem is $\max\{\tilde{\mathbf{f}}(\tilde{\mathbf{x}}) : \tilde{\mathbf{x}} \in \tilde{\mathbf{X}}\}$, **where** $\tilde{\mathbf{f}} : \Re^{\tilde{\mathbf{n}}} \to \Re$ **is defined as**

$$\tilde{\mathbf{f}}(\tilde{\mathbf{x}}) = -\tilde{\phi}(\tilde{\mathbf{x}}) + \min_{\mathbf{x} \in \mathbf{X}} \langle \mathbf{C}^* \tilde{\mathbf{x}}, \mathbf{x} \rangle + \phi(\mathbf{x}).$$

$\mathbf{f}$ **is non-smooth but it can be approximated by the smooth convex function**

$$\mathbf{f}_\mu(\mathbf{x}) = \phi(\mathbf{x}) + \max_{\tilde{\mathbf{x}} \in \tilde{\mathbf{X}}} \langle \tilde{\mathbf{x}}, \mathbf{C}\mathbf{x} \rangle - \tilde{\phi}(\tilde{\mathbf{x}}) - \mu\, \tilde{\mathbf{h}}(\tilde{\mathbf{x}}) \qquad (*)$$

where $\mu > \mathbf{0}$ **and** $\tilde{\mathbf{h}}$ **is a** $\tilde{\sigma}$**-strongly convex function such that** $\min\{\tilde{\mathbf{h}}(\tilde{\mathbf{x}}) : \tilde{\mathbf{x}} \in \tilde{\mathbf{X}}\} = \mathbf{0}$**.**

Proposition:

i) $(*)$ **has a unique optimal solution** $\tilde{\mathbf{x}}_\mu(\mathbf{x})$ **and** $\mathbf{f}'_\mu(\mathbf{x}) = \mathbf{C}^* \tilde{\mathbf{x}}_\mu(\mathbf{x})$**.**

ii) $\mathbf{f}'_\mu$ **is** $\mathbf{L}_\mu$**-Lips. cont. with** $\mathbf{L}_\mu := \mathbf{L}_\phi + \|\mathbf{C}\|^{\mathbf{2}}/(\mu\tilde{\sigma})$**.**

iii) $\mathbf{f}_\mu(\cdot) \le \mathbf{f}(\cdot) \le \mathbf{f}_\mu(\cdot) + \mu\tilde{\mathbf{D}}$, **where** $\tilde{\mathbf{D}} = \max_{\tilde{\mathbf{x}} \in \tilde{\mathbf{X}}} \tilde{\mathbf{h}}(\tilde{\mathbf{x}})$**.**

Nesterov's approximation scheme: Set $\mu = \epsilon/(2\tilde{D})$ and apply Nesterov's smooth method to f_μ with stepsize $\alpha_k = (k+1)/2$ until an iterate x_k s.t. $f(x_k) - f^* \le \epsilon$ is found.

Theorem (Nesterov): If $D \ge d_h(x^*; x_0)$ is known, then the above scheme generates a sequence $\{x_k\}$ satisfying

$$\begin{aligned} f(x_k) - f^* &\le \mu\tilde{D} + \left(L_\phi + \frac{\|C\|^2}{\mu\tilde{\sigma}}\right)\frac{4d_h(x^*; x_0)}{\sigma k(k+1)} \\ &\le \frac{\epsilon}{2} + \left(L_\phi + \frac{2\tilde{D}\|C\|^2}{\epsilon\tilde{\sigma}}\right)\frac{4D}{\sigma k^2} \quad (*) \end{aligned}$$

where $h : X \to \Re$ is the σ-strongly convex function used by Nesterov's smooth method. Hence, # of iterations of Nesterov's scheme is bounded by

$$\left\lceil \left(L_\phi + \frac{2\tilde{D}\|C\|^2}{\epsilon\tilde{\sigma}}\right)^{\frac{1}{2}} \left(\frac{8D}{\sigma\epsilon}\right)^{\frac{1}{2}} \right\rceil = \left\lceil \frac{4\|C\|}{\epsilon}\sqrt{\frac{D\tilde{D}}{\sigma\tilde{\sigma}}} \right\rceil$$

Lower bound result: Assume X is bounded and that now $D \ge \sup_{x \in X} d_h(x^*; x)$. Letting

$$\tilde{x}_k := \frac{\sum_{i=0}^k \alpha_i \tilde{x}_\mu(x_k)}{\sum_{i=0}^k \alpha_i}$$

then $\tilde{f}(\tilde{x}_k) \le f^*$ and $f(x_k) - \tilde{f}(\tilde{x}_k)$ is bounded by $(*)$.

Remarks:

1) Unless $\mathbf{f}^*$ is known or $\mathbf{X}$ is bounded, we do not know when $\mathbf{f}(\mathbf{x_k}) - \mathbf{f}^* \leq \epsilon$ occurs.

2) If $\tilde{\phi}(\tilde{\mathbf{x}})$ is already $\tilde{\sigma}$-strongly convex, then there is no need to add a perturbation term inside the inner maximization.

In this case, $\mathbf{f}$ is already $\mathbf{L_f}$-Lipschitz continuous with $\mathbf{L_f} = \mathbf{L}_\phi + \|\mathbf{C}\|^2/\tilde{\sigma}$, and we can apply Nesterov's optimal method directly to $\mathbf{f}$. The resulting complexity is

$$\mathbf{2}\|\mathbf{C}\|\sqrt{\frac{\mathbf{D}}{\sigma\tilde{\sigma}\epsilon}} = \mathcal{O}\left(\frac{\mathbf{1}}{\sqrt{\epsilon}}\right)$$

to find $\mathbf{x_k}$ such that $\mathbf{f}(\mathbf{x_k}) - \mathbf{f}^* \leq \epsilon$.

Example: Consider

$$\min_{\mathbf{x}} \mathbf{f}(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \|\mathbf{x}\|_1,$$

where $\mathbf{A}$ has full column rank. Can be reformulated as

$$\mathbf{f}(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \max\{\langle\, \mathbf{x}, \tilde{\mathbf{x}}\,\rangle : \|\tilde{\mathbf{x}}\|_\infty \leq \mathbf{1}\}$$

or as the dual $\max\{\tilde{\mathbf{f}}(\tilde{\mathbf{x}}) : \|\tilde{\mathbf{x}}\|_\infty \leq \mathbf{1}\}$, where

$$\begin{aligned}\tilde{\mathbf{f}}(\tilde{\mathbf{x}}) \quad &:= \quad \min_{\mathbf{x}} \langle\, \mathbf{x}, \tilde{\mathbf{x}}\,\rangle + \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \\ &= \quad -\max_{\mathbf{x}} \langle\, \mathbf{x}, -\tilde{\mathbf{x}}\,\rangle - \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2\end{aligned}$$

Variational Inequality: Assume

- $\mathbf{X} \subseteq \Re^{\mathbf{n}}$ is a non-empty closed convex set
- $\mathbf{F} : \mathbf{X} \to \Re^{\mathbf{n}}$ is a continuous monotone map

The (monotone) variational inequality (VI) problem $\mathbf{VIP}(\mathbf{F}, \mathbf{X})$ consists of finding $\mathbf{x}^*$ such that

$$\mathbf{x}^* \in \mathbf{X}, \qquad \min_{\mathbf{x} \in \mathbf{X}} \langle \mathbf{x} - \mathbf{x}^*, \mathbf{F}(\mathbf{x}) \rangle \geq \mathbf{0}$$

or equivalently,

$$\mathbf{x}^* \in \mathbf{X}, \qquad \min_{\mathbf{x} \in \mathbf{X}} \langle \mathbf{x} - \mathbf{x}^*, \mathbf{F}(\mathbf{x}^*) \rangle \geq \mathbf{0}$$

Assumptions:

- the set of solutions of $\mathbf{VIP}(\mathbf{F}, \mathbf{X})$ is nonempty.
- $\mathbf{F}$ is $\mathbf{L}$-Lipschitz continuous:

$$\|\mathbf{F}(\tilde{\mathbf{x}}) - \mathbf{F}(\mathbf{x})\| \leq \mathbf{L}\|\tilde{\mathbf{x}} - \mathbf{x}\|, \quad \forall \mathbf{x}, \, \tilde{\mathbf{x}} \in \mathbf{X}.$$

Remark: $\min\{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$ is clearly equivalent to $\mathbf{VI}(\mathbf{f}', \mathbf{X})$ and the second assumption is equivalent to $\mathbf{f}$ having $\mathbf{L}$-Lipschitz continuous gradient.

Saddle Point Problem

Assume that $\mathbf{\Phi} : \mathbf{U} \times \mathbf{V} \to \Re$ is a function such that

- $\mathbf{\Phi}(\cdot, \mathbf{v})$ is convex for all $\mathbf{v} \in \mathbf{V}$;
- $\mathbf{\Phi}(\mathbf{u}, \cdot)$ is concave for all $\mathbf{u} \in \mathbf{U}$.

Saddle point probl: Find $(\mathbf{u}^*, \mathbf{v}^*) \in \mathbf{U} \times \mathbf{V}$ such that

$$\mathbf{\Phi}(\mathbf{u}^*, \mathbf{v}) \leq \mathbf{\Phi}(\mathbf{u}^*, \mathbf{v}^*) \leq \mathbf{\Phi}(\mathbf{u}, \mathbf{v}^*), \quad \forall (\mathbf{u}, \mathbf{v}) \in \mathbf{U} \times \mathbf{V} \quad (*)$$

which, for $\mathbf{\Phi}$ smooth, is equivalent to $\mathbf{VI}(\mathbf{F}, \mathbf{X})$ with

$$\mathbf{F} = (\mathbf{\Phi}'_{\mathbf{u}}, -\mathbf{\Phi}'_{\mathbf{v}}), \quad \mathbf{X} = \mathbf{U} \times \mathbf{V}$$

Optimiz. view: Consider the pair of dual probl's

$$(\mathbf{P}) \quad \mathbf{f}^*_{\mathbf{P}} := \min_{\mathbf{u} \in \mathbf{U}} \left(\mathbf{f}_{\mathbf{P}}(\mathbf{u}) := \max_{\mathbf{v} \in \mathbf{V}} \mathbf{\Phi}(\mathbf{u}, \mathbf{v}) \right)$$

$$(\mathbf{D}) \quad \mathbf{f}^*_{\mathbf{D}} := \max_{\mathbf{v} \in \mathbf{V}} \left(\mathbf{f}_{\mathbf{D}}(\mathbf{v}) := \min_{\mathbf{u} \in \mathbf{U}} \mathbf{\Phi}(\mathbf{u}, \mathbf{v}) \right)$$

Clearly,

$$\mathbf{f}_{\mathbf{D}}(\mathbf{v}) \leq \mathbf{\Phi}(\mathbf{u}, \mathbf{v}) \leq \mathbf{f}_{\mathbf{P}}(\mathbf{u}), \quad \forall (\mathbf{u}, \mathbf{v}) \in \mathbf{U} \times \mathbf{V}$$

and hence $\mathbf{f}^*_{\mathbf{D}} \leq \mathbf{f}^*_{\mathbf{P}}$.

Proposition: For $(\mathbf{u}^*, \mathbf{v}^*) \in \mathbf{U} \times \mathbf{V}$, $(\mathbf{u}^*, \mathbf{v}^*)$ is a saddle point (i.e., satisfies $(*)$) if and only if $\mathbf{f}_{\mathbf{D}}(\mathbf{v}^*) = \mathbf{f}_{\mathbf{P}}(\mathbf{u}^*)$, in which case $\mathbf{f}^*_{\mathbf{D}} = \mathbf{\Phi}(\mathbf{u}^*, \mathbf{v}^*) = \mathbf{f}^*_{\mathbf{P}}$.

Algorithms for VI

Let $\mathbf{h} : \mathbf{X} \to \Re$ be a σ-strongly convex function and $\mathbf{d_h} : \mathbf{X} \times \mathbf{X} \to \Re$ be the associated Bregman distance:

$$\mathbf{d_h}(\tilde{\mathbf{x}}; \mathbf{x}) = \mathbf{h}(\tilde{\mathbf{x}}) - \mathbf{l_h}(\tilde{\mathbf{x}}; \mathbf{x}), \quad \forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathbf{X}$$

Prox-mirror method:

0) **Let $\mathbf{x_0} \in \mathbf{X}$ be given and set $\mathbf{k = 1}$.**

1) **Compute $\mathbf{F}(\mathbf{x_{k-1}})$ and let**

$$\mathbf{y_k} = \arg\min_{\mathbf{x} \in \mathbf{X}} \frac{\mathbf{1}}{\sqrt{\mathbf{2}}\mathbf{L}} \langle \mathbf{F}(\mathbf{x_{k-1}}), \mathbf{x} \rangle + \frac{\mathbf{1}}{\sigma} \mathbf{d_h}(\mathbf{x}; \mathbf{x_{k-1}})$$

2) **Compute $\mathbf{F}(\mathbf{y_k})$ and let**

$$\mathbf{x_k} = \arg\min_{\mathbf{x} \in \mathbf{X}} \frac{\mathbf{1}}{\sqrt{\mathbf{2}}\mathbf{L}} \langle \mathbf{F}(\mathbf{y_k}), \mathbf{x} \rangle + \frac{\mathbf{1}}{\sigma} \mathbf{d_h}(\mathbf{x}; \mathbf{x_{k-1}})$$

3) **Set $\mathbf{k} \leftarrow \mathbf{k + 1}$ and go to step 1.**

Remark: **When $\mathbf{h} = \|\cdot\|^{\mathbf{2}}/\mathbf{2}$, the above method reduces to Korpelevich's algorithm, i.e.:**

$$\begin{aligned} \mathbf{y_k} &= \mathbf{P_X}(\mathbf{x_{k-1}} - \lambda \mathbf{F}(\mathbf{x_{k-1}})), \\ \mathbf{x_k} &= \mathbf{P_X}(\mathbf{x_{k-1}} - \lambda \mathbf{F}(\mathbf{y_k})). \end{aligned}$$

where $\lambda := (\sqrt{\mathbf{2}}\mathbf{L})^{-\mathbf{1}}$.

Complexity results for VI

Theorem (Nemirovski): If $\mathbf{X}$ is bounded, then

$$\min_{\mathbf{x}\in\mathbf{X}} \langle \mathbf{F}(\mathbf{x}), \mathbf{x} - \bar{\mathbf{y}}_{\mathbf{k}} \rangle \geq -\frac{\sqrt{2}\,\mathbf{L}\,\mathbf{D}_{\mathbf{h}}(\mathbf{x}_0)}{\mathbf{k}\,\sigma}$$

for every $\mathbf{k} \geq 1$, where

$$\bar{\mathbf{y}}_{\mathbf{k}} := \frac{1}{\mathbf{k}}\sum_{\mathbf{i}=1}^{\mathbf{k}} \mathbf{y}_{\mathbf{i}}, \qquad \mathbf{D}_{\mathbf{h}}(\mathbf{x}_0) := \max_{\mathbf{x}\in\mathbf{X}} \mathbf{d}_{\mathbf{h}}(\mathbf{x}; \mathbf{x}_0)$$

Theorem (M. and Svaiter): If $\mathbf{h} = \|\cdot\|^2/2$, then there exist computable $(\bar{\mathbf{r}}_{\mathbf{k}}, \bar{\epsilon}_{\mathbf{k}}) \in \Re^{\mathbf{n}} \times \Re_+$ such that

$$\min_{\mathbf{x}\in\mathbf{X}} \langle \mathbf{F}(\mathbf{x}) - \bar{\mathbf{r}}_{\mathbf{k}}, \mathbf{x} - \bar{\mathbf{y}}_{\mathbf{k}} \rangle \geq -\bar{\epsilon}_{\mathbf{k}} \geq -\frac{2\sqrt{2}\,\mathbf{L}\,\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{\mathbf{k}\,\sigma}$$

$$\|\bar{\mathbf{r}}_{\mathbf{k}}\| \leq \frac{2\sqrt{2}\mathbf{L}\|\mathbf{x}_0 - \mathbf{x}^*\|}{\mathbf{k}}$$

If $\mathbf{F}$ is affine, then

$$\min_{\mathbf{x}\in\mathbf{X}} \langle \mathbf{F}(\bar{\mathbf{y}}_{\mathbf{k}}) - \bar{\mathbf{r}}_{\mathbf{k}}, \mathbf{x} - \bar{\mathbf{y}}_{\mathbf{k}} \rangle \geq -\bar{\epsilon}_{\mathbf{k}}$$

Remark: When $\mathbf{X} = \mathbf{K}$ is a cone, then the latter condition is equivalent to

$$\mathbf{F}(\bar{\mathbf{y}}_{\mathbf{k}}) - \bar{\mathbf{r}}_{\mathbf{k}} \in \mathbf{K}^*, \quad \langle \bar{\mathbf{y}}_{\mathbf{k}}, \mathbf{F}(\bar{\mathbf{y}}_{\mathbf{k}}) - \bar{\mathbf{r}}_{\mathbf{k}} \rangle \leq \bar{\epsilon}_{\mathbf{k}}$$

Complexity results for S.P. Problems

Let $\mathbf{F} = (\Phi'_{\mathbf{u}}, -\Phi'_{\mathbf{v}})$ and $\mathbf{X} = \mathbf{U} \times \mathbf{V}$. The S.P. problem is equivalent to $\mathbf{VIP}(\mathbf{F}, \mathbf{X})$.

Let $\{\bar{\mathbf{y}}_{\mathbf{k}}\} \subseteq \mathbf{U} \times \mathbf{V}$ be the ergodic sequence generated by the prox mirror method, and write $\bar{\mathbf{y}}_{\mathbf{k}} = (\bar{\mathbf{u}}_{\mathbf{k}}, \bar{\mathbf{v}}_{\mathbf{k}})$.

Theorem (Nemirovski): If $\mathbf{U} \times \mathbf{V}$ is bounded, then

$$\mathbf{f_P}(\bar{\mathbf{u}}_{\mathbf{k}}) - \mathbf{f_D}(\bar{\mathbf{v}}_{\mathbf{k}}) \leq \frac{\sqrt{2}\,\mathbf{L}\,\mathbf{D_h}(\mathbf{x_0})}{\mathbf{k}\,\sigma}$$

Theorem (M. and Svaiter): If $\mathbf{h} = \|\cdot\|^2/2$, then there exist computable $\bar{\mathbf{r}}_{\mathbf{k}} = (\bar{\mathbf{r}}^{\mathbf{u}}_{\mathbf{k}}, \bar{\mathbf{r}}^{\mathbf{v}}_{\mathbf{k}})$ and $\bar{\epsilon}_{\mathbf{k}} \geq \mathbf{0}$ such that the perturbed S.P. problem with

$$\Phi^{\mathbf{k}}(\mathbf{u}, \mathbf{v}) = \Phi(\mathbf{u}, \mathbf{v}) + \langle\, \bar{\mathbf{r}}^{\mathbf{u}}_{\mathbf{k}}, \mathbf{u}\,\rangle + \langle\, \bar{\mathbf{r}}^{\mathbf{v}}_{\mathbf{k}}, \mathbf{v}\,\rangle$$

satisfies

$$\mathbf{f}^{\mathbf{k}}_{\mathbf{P}}(\bar{\mathbf{u}}_{\mathbf{k}}) - \mathbf{f}^{\mathbf{k}}_{\mathbf{D}}(\bar{\mathbf{v}}_{\mathbf{k}}) \leq \bar{\epsilon}_{\mathbf{k}} \leq \frac{2\sqrt{2}\,\mathbf{L}\,\|\mathbf{x_0} - \mathbf{x}^*\|^2}{\mathbf{k}\,\sigma}$$

$$\|\bar{\mathbf{r}}_{\mathbf{k}}\| \leq \frac{2\sqrt{2}\mathbf{L}\|\mathbf{x_0} - \mathbf{x}^*\|}{\mathbf{k}}$$

where $\mathbf{f}^{\mathbf{k}}_{\mathbf{P}}$ and $\mathbf{f}^{\mathbf{k}}_{\mathbf{D}}$ are the associated primal and dual functions.

THANK YOU!

AND

THE END

Relative scale bounds

Consider the problem $\mathbf{f^* = \min\{f(x) : Cx = d\}}$, where $\mathbf{C}$ is $\mathbf{m \times n}$, $\mathbf{0 \neq d \in \Re^m}$ and $\mathbf{f : \Re^n \to \Re}$ is convex, homogenous of degree 1 and $\mathbf{0 \in \operatorname{int} \partial f(0)}$.

For some inner product norm $\|\cdot\|$, assume that

$$\mathbf{B(0; m) \subseteq \partial f(0) \subseteq B(0; M)}$$

for some $\mathbf{0 < m \leq M}$, or equivalently, $\mathbf{m\|x\| \leq f(x) \leq M\|x\|}$ for all $\mathbf{x}$. Clearly, $\mathbf{f^* > 0}$.

Lemma: Let $\mathbf{x_0 := \operatorname{argmin}\{\|x\| : Cx = d\}}$. Then,

$$\mathbf{R := \frac{f(x^0)}{m} \geq \frac{f^*}{m} \geq \|x_0 - x^*\|}, \qquad \mathbf{\frac{f(x^0)}{f^*} \leq \frac{M}{m}}$$

Proposition (Nesterov): The subgradient method with stepsize $\mathbf{\alpha_k = R/\sqrt{K+1}}$, $\mathbf{k = 0, \ldots, K}$, where $\mathbf{R}$ is as above and

$$\mathbf{K := \left\lfloor \frac{(M/m)^4}{\delta^2} \right\rfloor}$$

satisfies

$$\mathbf{\frac{1}{f^*}\left(\min_{i=0,\ldots,K} [f(x_i) - f^*]\right) \leq \delta}$$

Remark: The norm should be chosen so that $\mathbf{M/m}$ is as small as possible.