Math 273a: Optimization
Convex Conjugacy

Instructor: Wotao Yin
Department of Mathematics, UCLA

Fall 2015

online discussions on piazza.com

Convex conjugate (the Legendre transform)

Let f be a closed proper convex function.

The convex conjugate of f is

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbf{dom} f} \{\mathbf{y}^T\mathbf{x} - f(\mathbf{x})\}$$

- f^* is convex (in $\mathbf{y}$).

Reason: for each fixed $\mathbf{x}$, $(\mathbf{y}^T\mathbf{x} - f(\mathbf{x}))$ is linear in $\mathbf{y}$. Hence, f^* is point-wise maximum of linear functions, that is, point $\mathbf{y}$ and over $\mathbf{x}$.

- As long as f is proper, f^* is proper closed convex.

Geometry

- For fixed $\mathbf{y}$ and z, consider linear function: $g(\mathbf{x}) = \mathbf{y}^T\mathbf{x} - z$
  - the corresponding hyperplane is

$$\mathcal{H} = \{(\mathbf{x}, g(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$$

  - $[\mathbf{y}; -1]$ is the *downward normal direction* of $\mathcal{H}$
  - $\mathcal{H}$ crosses the $(n+1)$th axis at $g(\mathbf{0}) = -z$
  - $\mathbf{y}$ (tilt) and z (height) uniquely define the hyperplane $\mathcal{H}$ and function g

Our task: for each $\mathbf{y}$, determine z using the function f, thus determining g

- If we set **two rules**

1. $g(\mathbf{x}) = f(\mathbf{x})$ at some point $\mathbf{x}$, i.e., $\mathcal{H}$ intersects $\mathrm{epi}(f)$,
2. $\mathcal{H}$ is *as low as possible*, i.e., $-z$ is *as small as possible*,

then $\mathcal{H}$ will be the *supporting hyperplane* of f

By rule #1, $\exists$ point $\mathbf{x} \in \mathbf{dom} f$, $\ni \mathbf{y}^T\mathbf{x} - z = f(\mathbf{x})$ or $-z = f(\mathbf{x}) - \mathbf{y}^T\mathbf{x}$
By rule #2, $-z = \inf_{\mathbf{dom} f}\{f(\mathbf{x}) - \mathbf{y}^T\mathbf{x}\} \implies z = \sup_{\mathbf{x}\in\mathbf{dom} f}\{\mathbf{y}^T\mathbf{x} - f(\mathbf{x})\}$

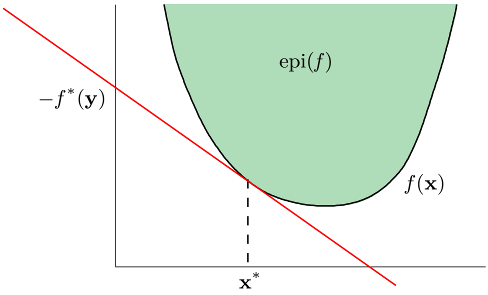
Therefore,

- $z = f^*(\mathbf{x})$
- $g(\mathbf{y}) = \mathbf{y}^T\mathbf{x} - f^*(\mathbf{x})$
- $\mathcal{H}$ is the supporting hyperplane

Geometry

1. $\mathcal{H}$ intersects $\mathrm{epi}(f)$
2. $\mathcal{H}$ is *as low as possible*

$\implies -f^*(\mathbf{y}) = \inf_{\mathbf{x} \in \mathbf{dom} f} \{f(\mathbf{x}) - \mathbf{y}^T \mathbf{x}\}$



f^* almost completely characterizes f. "Almost" is because covers only up to closure. This is a result of the Hahn-Banach Separation Theorem.

Relation to Lagrange duality

Consider convex problem

$$\underset{\mathbf{x}}{\text{minimize}}\, f(\mathbf{x}) \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

Lagrangian:

$$\mathcal{L}(\mathbf{x}; \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

Lagrange dual function:

$$\begin{aligned} \mathbf{d}(\mathbf{y}) &= -\inf_{\mathbf{x} \in \mathbf{dom} f} \mathcal{L}(\mathbf{x}; \mathbf{y}) \\ &= \sup_{\mathbf{x} \in \mathbf{dom} f} \{\mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) - f(\mathbf{x})\} \\ &= f^*(\mathbf{A}^T\mathbf{y}) - \mathbf{b}^T\mathbf{y}. \end{aligned}$$

Lagrange dual problem (given in terms of convex conjugate f^*):

$$\underset{\mathbf{y}}{\text{minimize}}\, f^*(\mathbf{A}^T\mathbf{y}) - \mathbf{b}^T\mathbf{y} \quad \text{or} \quad \underset{\mathbf{y}}{\text{maximize}}\, \mathbf{b}^T\mathbf{y} - f^*(\mathbf{A}^T\mathbf{y}).$$

Exercise

Derive a Lagrange dual problem for

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; f(\mathbf{x}) + g(\mathbf{A}\mathbf{x})$$

Primal and dual subdifferentials

Suppose that $[\mathbf{y}; -1]$ is the *downward normal* of the hyperplane touching $\mathrm{epi}(f)$ at $\mathbf{x}$; therefore,

$$\mathbf{y} = \nabla f(\mathbf{x}^*)$$

In general, for proper closed convex f,

$$\mathbf{y} \in \partial f(\mathbf{x}^*).$$

Therefore,

$$\mathbf{dom} f^* = \{\partial f(\mathbf{x}) : \mathbf{x} \in \mathbf{dom} f\}.$$

Theorem (biconjugation)

Let $\mathbf{x} \in \mathbf{dom} f$ *and* $\mathbf{y} \in \mathbf{dom} f^*$. *Then,*

$$\mathbf{y} \in \partial f(\mathbf{x}) \iff \mathbf{x} \in \partial f^*(\mathbf{y}).$$

If the relation holds, then

$$f(\mathbf{x}) + f^*(\mathbf{y}) = \mathbf{y}^T \mathbf{x}.$$

The result is very useful in deriving optimality conditions.

Fenchel's inequality

Theorem (Fenchel's inequality)

For arbitrary $\mathbf{x} \in \mathbf{dom} f$ *and* $\mathbf{y} \in \mathbf{dom} f^*$, *we have*

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \mathbf{y}^T \mathbf{x}.$$

Proof.

Since $\mathbf{x}$ is not necessarily the maximizing point for $f(\mathbf{y}) = \sup_{\mathbf{x}}\{\cdots\}$, we have

$$f^*(\mathbf{y}) \geq \mathbf{y}^T \mathbf{x} - f(\mathbf{x}).$$

□

Theorem

If f is proper, closed, convex, then $(f^*)^* = f$, i.e.,

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} f^*} \{\mathbf{y}^T \mathbf{x} - f^*(\mathbf{y})\}.$$

Proof.

Consider linear function $g_{\mathbf{y},z}$, defined as $g_{\mathbf{y},z}(\mathbf{x}) = \mathbf{y}^T \mathbf{x} - z$.

Step 1.

$$\begin{aligned} g_{\mathbf{y},z} \le f &\iff \mathbf{y}^T \mathbf{x} - z \le f(\mathbf{x}), \ \forall \mathbf{x} \\ &\iff \mathbf{y}^T \mathbf{x} - f(\mathbf{x}) \le z, \ \forall \mathbf{x} \\ &\iff \sup_{\mathbf{x}} \{\mathbf{y}^T \mathbf{x} - f(\mathbf{x})\} \le z \\ &\iff f^*(\mathbf{y}) \le z \\ &\iff (\mathbf{y}, z) \in \mathrm{epi}(f^*) \end{aligned}$$

(cont.) □

Proof.

Step 2. From the Hahn-Banach Separation Theorem,

$$f(\mathbf{x}) = \sup_{\mathbf{y},z}\{g_{\mathbf{y},z}(\mathbf{x}) : g_{\mathbf{y},z} \le f\}, \quad \forall \mathbf{x} \in \mathbf{dom} f.$$

Step 3.

$$\begin{aligned}
\sup_{\mathbf{y},z}\{g_{\mathbf{y},z}(\mathbf{x}) : g_{\mathbf{y},z} \le f\} &= \sup_{\mathbf{y},z}\{g_{\mathbf{y},z}(\mathbf{x}) : f^*(\mathbf{y}) \le z\} \text{ by Step 1} \\
&= \sup_{\mathbf{y},z}\{\mathbf{y}^T\mathbf{x} - z : f^*(\mathbf{y}) \le z\} \\
&= \sup_{\mathbf{y}}\{\mathbf{y}^T\mathbf{x} - f^*(\mathbf{y})\} \\
&= (f^*)^*(\mathbf{x})
\end{aligned}$$

Combining Steps 2 and 3, we get $f = (f^*)^*$. □

Examples

- $f(\mathbf{x}) = \iota_{\mathcal{C}}(\mathbf{x})$, *indicator function* of nonempty closed convex set $\mathcal{C}$; then

$$\sigma_{\mathcal{C}}^* := f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathcal{C}} \mathbf{y}^T \mathbf{x}$$

  is the *support function* of $\mathcal{C}$

- Applying the theorem, we get $(\iota_{\mathcal{C}}^*)^* = \iota_{\mathcal{C}}$ and $(\sigma_{\mathcal{C}}^*)^* = \sigma_{\mathcal{C}}$.

Examples

- $f(\mathbf{x}) = \iota_{\{-1 \le \mathbf{x} \le 1\}}$, indicator function of the unit hypercube; then

$$f^*(\mathbf{y}) = \sup_{-1 \le \mathbf{x} \le 1} \mathbf{y}^T \mathbf{x} = \|\mathbf{y}\|_1$$

- $f(\mathbf{x}) = \iota_{\{\|\mathbf{x}\|_2 \le 1\}}$, then

$$f^*(\mathbf{y}) = \|\mathbf{y}\|_2$$

- $f(\mathbf{x}) = \frac{1}{p}\|\mathbf{x}\|_p^p$, $1 < p < \infty$, then

$$f^*(\mathbf{y}) = \frac{1}{q}\|\mathbf{x}\|_q^q$$

  $\frac{1}{p} + \frac{1}{q} = 1$

- lots of smooth examples ......

Alternative representation

Previously, we can represent f by f^* via

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} f^*} \{\mathbf{x}^T \mathbf{y} - f^*(\mathbf{y})\}$$

We can introduce a more general representation:

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} h^*} \{(\mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{y} - h^*(\mathbf{y})\} = h(\mathbf{A}\mathbf{x} - \mathbf{b})$$

so that h^* might be simpler than f^* (or h is simpler than f) in form.

Example $f(\mathbf{x}) = \|\mathbf{x}\|_1$

- Let $\mathcal{C} = \{\mathbf{y} = [\mathbf{y}_1; \mathbf{y}_2] : \mathbf{y}_1 + \mathbf{y}_2 = \mathbf{1},\ \mathbf{y}_1, \mathbf{y}_2 \geq 0\}$ and

$$\mathbf{A} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

- We have

$$\|\mathbf{x}\|_1 = \sup_{\mathbf{y}} \{(\mathbf{y}_1 - \mathbf{y}_2)^T \mathbf{x} - \iota_{\mathcal{C}}(\mathbf{y})\} = \sup_{\mathbf{y}} \{\mathbf{y}^T \mathbf{A}\mathbf{x} - \iota_{\mathcal{C}}(\mathbf{y})\}.$$

- Since $\iota_{\mathcal{C}}^*([\mathbf{x}_1; \mathbf{x}_2]) = \mathbf{1}^T (\max\{\mathbf{x}_1, \mathbf{x}_2\})$, where $\max$ is taken entry-wise, and

$$\sup_{\mathbf{y} \in \mathbf{dom} h^*} \{(\mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{y} - h^*(\mathbf{y})\} = h(\mathbf{A}\mathbf{x} - \mathbf{b})$$

  we have

$$\|\mathbf{x}\|_1 = \iota_{\mathcal{C}}^*(\mathbf{A}\mathbf{x}) = \mathbf{1}^T (\max\{\mathbf{x}, -\mathbf{x}\}).$$

Application: dual smoothing

Idea: "strongly convexify" h^* $\Longrightarrow$ f becomes Lipschitz-differentiable

- Suppose that f is represented in terms of h^* as

$$f(\mathbf{x}) = \sup_{\mathbf{y}\in\mathbf{dom}h^*} \{\mathbf{y}^T(\mathbf{A}\mathbf{x}-\mathbf{b}) - h^*(\mathbf{y})\}$$

- Let us strongly convexify h^* by adding strongly convex function d:

$$\hat{h}^*(\mathbf{y}) = h^*(\mathbf{y}) + \mu d(\mathbf{y})$$

  (a simple choice is $d(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2$)

- Obtain a Lipschitz-differentiable approximation:

$$f_\mu(\mathbf{x}) = \sup_{\mathbf{y}\in\mathbf{dom}h^*} \{\mathbf{y}^T(\mathbf{A}\mathbf{x}-\mathbf{b}) - \hat{h}^*(\mathbf{y})\}$$

- $f_\mu(\mathbf{x})$ is differentiable since $h^*(\mathbf{y}) + \mu d(\mathbf{y})$ is strongly convex.

Example: augmented ℓ_1

- primal problem: $\min\{\|\mathbf{x}\|_1 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$
- dual problem: $\max\{\mathbf{b}^T\mathbf{y} + \iota_{[-1,1]^n}(\mathbf{A}^T\mathbf{y})\}$
- $f(\mathbf{y}) = \iota_{[-1,1]^n}(\mathbf{y})$ is non-differentiable

  plan: strongly convexify the primal so that dual becomes smooth and can be quickly solved
- let $f^*(\mathbf{x}) = \|\mathbf{x}\|_1$ and $f(\mathbf{y}) = \iota_{[-1,1]^n}(\mathbf{y}) = \sup_{\mathbf{x}}\{\mathbf{y}^T\mathbf{x} - f^*(\mathbf{x})\}$,
- add $\frac{\mu}{2}\|\mathbf{x}\|_2$ to $f^*(\mathbf{x})$ and obtain

$$f_\mu(\mathbf{y}) = \sup_{\mathbf{x}}\{\mathbf{y}^T\mathbf{x} - (\|\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{x}\|_2^2)\} = \frac{1}{2\mu}\|\mathbf{y} - \mathrm{Proj}_{[-1,1]^n}(\mathbf{y})\|_2^2$$

- $f_\mu(\mathbf{y})$ is differentiable; $\nabla f_\mu(\mathbf{y}) = \frac{1}{\mu}\,\mathrm{shrink}(\mathbf{y})$.
- On the other hand, we can also directly smooth $f^*(\mathbf{x}) = \|\mathbf{x}\|_1$ and obtain differentiable $f_\mu^*(\mathbf{x})$ by adding $d(\mathbf{y})$ to $f(\mathbf{y})$. (see the next slide ...)

Example: smoothed absolute value

Let $x \in \mathbb{R}$. Recall

$$f(x) = |x| = \sup_y \{yx - \iota_{[-1,1]}(y)\}$$

- add $d(y) = y^2/2$ to $\iota_{[-1,1]}(y)$ and obtain:

$$f_\mu = \sup_y \{yx - (\iota_{[-1,1]}(y) + \mu y^2/2)\} = \begin{cases} x^2/(2\mu), & |x| \le \mu, \\ |x| - \mu/2, & |x| > \mu, \end{cases}$$

which is the **Huber function**.

- If $|x| \le \mu$, the maximizing y stays within $[-1, 1]$ even if $\iota_{[-1,1]}(y)$ vanishes, so it leaves with $\mu y^2/2$ only.
- If $|x| > \mu$, the maximizing y occurs at ± 1, so $-\mu y^2/2 = -\mu/2$.
- The Huber function is used in robust least squares.

- add $d(y) = 1 - \sqrt{1-y^2}$, which is well defined and strongly convex in $[-1,1]$:

$$f_\mu^* = \sup_y \{yx - (\iota_{[-1,1]}(y) - \mu\sqrt{1-y^2})\} - \mu = \sqrt{x^2+\mu^2} - \mu,$$

which is used in reweighted least-squares methods

- Recall

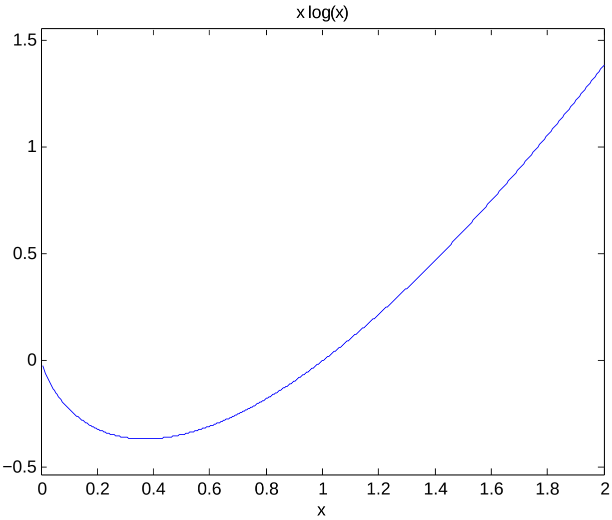
$$|x| = \sup_{\mathbf{y}} \{(y_1 - y_2)x - \iota_{\mathcal{C}}(y)\}$$

for $\mathcal{C} = \{\mathbf{y} : y_1 + y_2 = 1,\ y_1, y_2 \geq 0\}$.

Add negative entropy $d(y) = y_1 \log y_1 + y_2 \log y_2 + \log 2$:

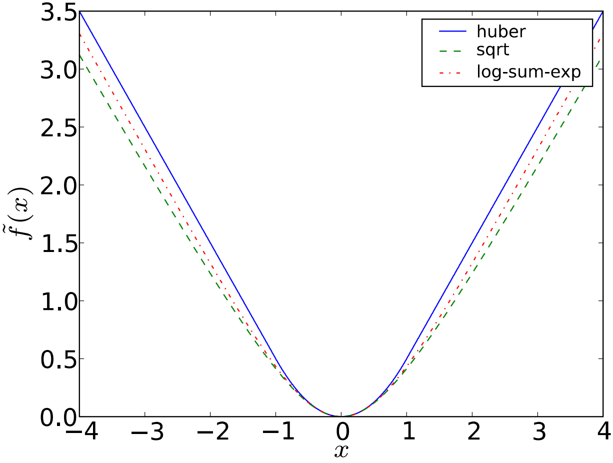
$$f_\mu^*(x) = \sup_{\mathbf{y}} \{(y_1 - y_2)x - (\iota_{\mathcal{C}}(y) + \mu d(y))\} = \mu \log \frac{e^{x/\mu} + e^{-x/\mu}}{2}.$$

$x \log(x)$



$x \log(x)$ is strongly convex between $[0, C]$ for any finite $C > 0$

Compare three smoothed functions



Courtesy of L. Vandenberghe

Example: smoothed maximum eigenvalue

- Let $\mathbf{X} \in \mathcal{S}^n$. Consider

$$f(\mathbf{X}) = \lambda_{\max}(\mathbf{X}),$$

which is the "ℓ_∞ norm" on the matrix spectrum.

- Recall the dual of ℓ_∞ is ℓ_1, and we have

$$\|\mathbf{x}\|_\infty = \sup_{\mathbf{y}} \{\mathbf{x}^T\mathbf{y} - \iota_{\{\mathbf{1}^T\mathbf{y}=1, \mathbf{y}\geq 0\}}(\mathbf{y})\}$$

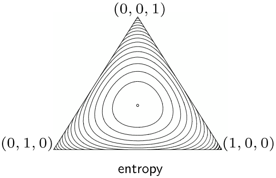
- Let $\mathcal{C} = \{\mathbf{Y} \in \mathcal{S}^n : \mathrm{tr}\mathbf{Y} = 1, \mathbf{Y} \succeq 0\}$. We have

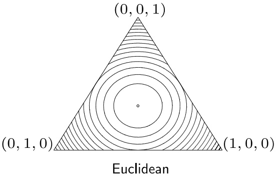
$$f(\mathbf{X}) = \lambda_{\max}(\mathbf{X}) = \sup_{\mathbf{Y}} \{\mathbf{Y} \bullet \mathbf{X} - \iota_{\mathcal{C}}(\mathbf{Y})\}$$

- Next, strongly convexify $\iota_{\mathcal{C}}(\mathbf{Y})$:

Negative entropy of $\{\lambda_i(\mathbf{Y})\}$:

$$d(\mathbf{Y}) = \sum_{i=1}^{n} \lambda_i(\mathbf{Y}) \log \lambda_i(\mathbf{Y}) + \log n$$





(Courtesy of L. Vandenberghe)

Smoothed function

$$f_\mu(\mathbf{X}) = \sup_{\mathbf{Y}} \{\mathbf{Y} \bullet \mathbf{X} - (\iota_{\mathcal{C}}(\mathbf{Y}) + \mu d(\mathbf{Y}))\} = \mu \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{\lambda_i(\mathbf{X})/\mu} \right)$$

Application: smoothed minimization[1]

Instead of solving

$$\min_{\mathbf{x}} f(\mathbf{x}),$$

solve

$$\min_{\mathbf{x}} f_\mu(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} h^*} \left\{ \mathbf{y}^T(\mathbf{A}\mathbf{x} + \mathbf{b}) - [h^*(\mathbf{y}) + \mu d(\mathbf{y})] \right\}$$

by gradient descent, with acceleration, line search, etc......

[1] ?

Since $h^*(\mathbf{y}) + \mu d(\mathbf{y})$ is strongly convex, $\nabla f_\mu(\mathbf{x})$ is given by:

$$\nabla f_\mu(\mathbf{x}) = \mathbf{A}^T \bar{\mathbf{y}}, \quad \text{where } \bar{\mathbf{y}} = \arg\max_{\mathbf{y} \in \mathbf{dom} h^*} \{\mathbf{y}^T(\mathbf{A}\mathbf{x} + \mathbf{b}) - [h^*(\mathbf{y}) + \mu d(\mathbf{y})]\}.$$

Theorem

If $d(\mathbf{y})$ is strongly convex with modulus $\nu > 0$, then

- $h^*(\mathbf{y}) + \mu d(\mathbf{y})$ *is strongly convex with modulus at least* $\mu\nu$
- $\nabla f_\mu(\mathbf{x})$ *is Lipschitz continuous with constant no more than* $\|\mathbf{A}\|^2/\mu\nu$.

Nonsmooth optimization

Examples:

$$\min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1$$

$$\min \mathrm{TV}(\mathbf{x}) \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \sigma$$

Worst-case complexity for ϵ-approximation:

- If f is convex and ∇f is L-Lipschitz, accelerated gradient method takes

  $$O(\sqrt{L/\epsilon})$$

  iterations.

- If f is convex and nonsmooth, f is G-Lipschitz, subgradient method takes

  $$O(G^2/\epsilon^2)$$

  iterations.

Smooth optimization has much better complexities.

Nesterov's complexity analysis

1. Construct smooth approximate satisfying

$$f_\mu \leq f \leq f_\mu + \mu D$$

   and consequently

$$f(\mathbf{x}) - f^* \leq f_\mu(x) - f_\mu^* + \mu D$$

2. Choose μ such that $\mu D \leq \epsilon/2 \implies \frac{1}{\mu} \geq \frac{2D}{\epsilon}$
3. Minimize f_μ such that $f_\mu(x) - f_\mu^* \leq \epsilon/2$

Step 3 has complexity

$$O(\sqrt{\frac{1}{\mu\epsilon}}) = O(\frac{\sqrt{D}}{\epsilon}),$$

which can be much better than the subgradient method's

$$O(G^2/\epsilon^2).$$

Theorem

Consider

$$h(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} h^*} \mathbf{y}^T \mathbf{x} - h^*(\mathbf{y}),$$

$$h_\mu(\mathbf{x}) = \sup_{\mathbf{y} \in \mathbf{dom} h^*} \mathbf{y}^T \mathbf{x} - (h^*(\mathbf{y}) + \mu d(\mathbf{y})),$$

where $d(\mathbf{y}) \geq 0$ is strongly convex with modulus $\nu > 0$. Then,

1. ∇h_μ *is* $(\mu\nu)^{-1}$*-Lipschitz;*
2. *if* $d(\mathbf{y}) \leq D$ *for* $\mathbf{y} \in \mathbf{dom} h^*$*, then*

$$h_\mu(\mathbf{x}) \leq h(\mathbf{x}) \leq h_\mu(\mathbf{x}) + \mu D, \quad \mathbf{x} \in \mathbf{dom} h.$$

Example: Huber function

Recall

$$h_\mu(x) = \sup_y \{yx - (\iota_{[-1,1]}(y) + \mu y^2/2)\} = \begin{cases} x^2/(2\mu), & |x| \le \mu, \\ |x| - \mu/2, & |x| > \mu, \end{cases}$$

We have

$$h_\mu(x) \le |x| \le h_\mu(x) + \mu/2$$

and

$$h'_\mu(x) = \begin{cases} x/\mu, & |x| \le \mu, \\ \mathrm{sign}(x), & |x| > \mu, \end{cases}$$

which is Lipschitz with constant μ^{-1}.

Apply to ℓ_1-norm:

$$\sum_i h_\mu(x_i) \le \|\mathbf{x}\|_1 \le \sum_i h_\mu(x_i) + n\mu/2.$$

Robust least squares

Consider

$$\min_{\mathbf{x}} f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1$$

- Representation

$$\|\mathbf{x}\|_1 = \sup_{\mathbf{y}}\{\mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) - \iota_{\mathcal{C}}(\mathbf{y})\}$$

  where $\mathcal{C} = \{\mathbf{y} : \|\mathbf{y}\|_\infty \leq 1\}$.

- Add $\mu d(\mathbf{y}) = \frac{\mu}{2}\|\mathbf{y}\|_2^2$ to $\iota_{\mathcal{C}}(\mathbf{y})$ and obtain

$$\min_{\mathbf{x}} f_\mu(\mathbf{x}) = \sum_{i=1}^{m} h_\mu(\mathbf{a}_i^T\mathbf{x} - b_i).$$

Other examples and questions

- Total variation / analysis ℓ_1 minimization examples
- Nuclear norm examples
- More other one nonsmooth terms
- ...
- Stopping criteria