8. Proximal point method

- proximal point method
- augmented Lagrangian method
- Moreau–Yosida smoothing
Proximal point method

a “conceptual” algorithm for minimizing a closed convex function $f$:

$$x_{k+1} = \text{prox}_{t_k f}(x_k) = \arg\min_u \left( f(u) + \frac{1}{2t_k} \|u - x_k\|^2 \right)$$

- can be viewed as proximal gradient method (page 4.3) with $g(x) = 0$
- of interest if prox evaluations are much easier than minimizing $f$ directly
- a practical algorithm if inexact prox evaluations are used
- step size $t_k > 0$ affects number of iterations, cost of prox evaluations
- basis of the augmented Lagrangian method
Convergence

Assumptions

• \( f \) is closed and convex (hence, \( \text{prox}_{tf}(x) \) is uniquely defined for all \( x \))

• optimal value \( f^* \) is finite and attained at \( x^* \)

Result

\[
f(x_k) - f^* \leq \frac{\|x_0 - x^*\|_2^2}{2 \sum_{i=0}^{k-1} t_i}
\]

for \( k \geq 1 \)

• implies convergence if \( \sum_i t_i \rightarrow \infty \)

• rate is \( 1/k \) if \( t_i \) is fixed, or variable but bounded away from zero

• \( t_i \) is arbitrary; however cost of \( \text{prox} \) evaluations will depend on \( t_i \)
Proof: apply analysis of proximal gradient method (lecture 4) with $g(x) = 0$

- since $g$ is zero, inequality (3) on page 4.13 holds for any $t > 0$

- from page 4.15, $f(x_i)$ is nonincreasing and

$$t_i \left( f(x_{i+1}) - f^* \right) \leq \frac{1}{2} \left( \|x_i - x^*\|^2_2 - \|x_{i+1} - x^*\|^2_2 \right)$$

- combine inequalities for $i = 0$ to $i = k - 1$ to get

$$\left( \sum_{i=0}^{k-1} t_i \right) \left( f(x_k) - f^* \right) \leq \sum_{i=0}^{k-1} t_i \left( f(x_i) - f^* \right) \leq \frac{1}{2} \|x_0 - x^*\|^2_2$$

Proximal point method 8.4
Accelerated proximal point algorithms

- we take \( g(x) = 0 \) in FISTA on page 7.8:

\[
x_1 = \text{prox}_{t_0 f}(x_0)
\]

\[
x_{k+1} = \text{prox}_{t_k f} \left( x_k + \theta_k \left( \frac{1}{\theta_{k-1}} - 1 \right) (x_k - x_{k-1}) \right) \quad \text{for } k \geq 1
\]

- choose any \( t_k > 0 \), determine \( \theta_k \) from equation

\[
\frac{\theta_k^2}{t_k} = \left( 1 - \theta_k \right) \frac{\theta_{k-1}^2}{t_{k-1}}
\]

- converges if \( \sum_i \sqrt{t_i} \rightarrow \infty \) (lecture 7)

- rate is \( 1/k^2 \) if \( t_i \) is fixed or variable but bounded away from zero
Outline

• proximal point method

• **augmented Lagrangian method**

• Moreau–Yosida smoothing
Standard problem format

Primal and dual problem (page 5.21)

primal: minimize $f(x) + g(Ax)$

dual: maximize $-g^*(z) - f^*(-A^T z)$

Examples

• set constraints ($g(y) = \delta_C(y)$):

  minimize $f(x)$
  subject to $Ax \in C$

• regularized norm approximation ($g(y) = \|y - b\|$):

  minimize $f(x) + \|Ax - b\|$

Augmented Lagrangian method: proximal point method applied to the dual
Proximal mapping of dual function

**Definition:** proximal mapping of \( h(z) = f^*(-A^T z) + g^*(z) \) is defined as

\[
\text{prox}_{th}(z) = \arg\min_u \left( f^*(-A^T u) + g^*(u) + \frac{1}{2t} \|u - z\|^2 \right)
\]

**Dual expression:** \( \text{prox}_{th}(z) = z + t(A\hat{x} - \hat{y}) \) where

\[
(\hat{x}, \hat{y}) = \arg\min_{x, y} \left( f(x) + g(y) + z^T(Ax - y) + \frac{t}{2} \|Ax - y\|^2 \right)
\]

\( \hat{x}, \hat{y} \) minimize the *augmented Lagrangian* (Lagrangian + quadratic penalty)
Proof.

• write augmented Lagrangian minimization as

\[
\begin{align*}
\text{minimize (over } x, y, w) & \quad f(x) + g(y) + \frac{t}{2} \|w\|_2^2 \\
\text{subject to} & \quad Ax - y + \frac{z}{t} = w
\end{align*}
\]

• optimality conditions (\(u\) is multiplier for equality):

\[
Ax - y + \frac{1}{t}z = w, \quad -A^T u \in \partial f(x), \quad u \in \partial g(y), \quad tw = u
\]

• eliminating \(x, y, w\) gives \(u = z + t(Ax - y)\) and

\[
0 \in -A\partial f^*(-A^T u) + \partial g^*(u) + \frac{1}{t}(u - z)
\]

this is the optimality condition for problem in the definition of \(u = \text{prox}_{th}(z)\)
Augmented Lagrangian method

choose initial $z_0$ and repeat:

1. minimize augmented Lagrangian

\[(\hat{x}, \hat{y}) = \arg\min_{x,y} \left( f(x) + g(y) + \frac{t_k}{2} \| Ax - y + (1/t_k)z_k \|_2^2 \right)\]

2. dual update

\[z_{k+1} = z_k + t_k (A\hat{x} - \hat{y})\]

- also known as method of multipliers
- this is the proximal point method applied to the dual problem
- as variants, can apply the accelerated proximal point methods to the dual
- usually implemented with inexact minimization in step 1
Examples

minimize \( f(x) + g(Ax) \)

Equality constraints \((g\) is indicator of \(\{b\}\)):  
\[
\hat{x} = \arg\min_x \left( f(x) + z^T A x + \frac{t}{2} \| A x - b \|_2^2 \right) \\
z := z + t(A\hat{x} - b)
\]

Set constraint \((g\) indicator of convex set \(C\)):  
\[
\hat{x} = \arg\min_x \left( f(x) + \frac{t}{2} d(Ax + z/t)^2 \right) \\
z := z + t(A\hat{x} - PC(A\hat{x} + z/t))
\]

\(PC(u)\) is projection of \(u\) on \(C\), \(d(u) = \|u - PC(u)\|_2\) is Euclidean distance
Outline

• proximal point method

• augmented Lagrangian method

• Moreau–Yosida smoothing
Moreau–Yosida smoothing

**Definition:** the Moreau–Yosida regularization of a closed convex function $f$ is

$$f(t)(x) = \inf_u \left( f(u) + \frac{1}{2t} \|u - x\|^2 \right) \quad \text{(with } t > 0)$$

$$= f \left( \text{prox}_{tf}(x) \right) + \frac{1}{2t} \|\text{prox}_{tf}(x) - x\|^2$$

this is also known as the *Moreau envelope* of $f$

**Immediate properties**

- $f(t)$ is convex (infimum over $u$ of a convex function of $x$, $u$)
- domain of $f(t)$ is $\mathbb{R}^n$ (recall that $\text{prox}_{tf}(x)$ is defined for all $x$)
Examples

**Indicator function:** smoothed $f$ is squared Euclidean distance

$$f(x) = \delta_C(x), \quad f_t(x) = \frac{1}{2t} d(x)^2$$

**1-norm:** smoothed function is Huber penalty

$$f(x) = \|x\|_1, \quad f_t(x) = \sum_{k=1}^{n} \phi_t(x_k)$$

$$\phi_t(z) = \begin{cases} 
  \frac{z^2}{2t} & |z| \leq t \\
  |z| - t/2 & |z| \geq t
\end{cases}$$
Conjugate of Moreau envelope

\[ f_t(x) = \inf_u \left( f(u) + \frac{1}{2t} \| u - x \|^2 \right) \]

- \( f_t \) is infimal convolution of \( f(u) \) and \( \| v \|^2 / (2t) \) (see page 5.11):

\[ f_t(x) = \inf_{u+v=x} \left( f(u) + \frac{1}{2t} \| v \|^2 \right) \]

- from page 5.11, conjugate is sum of conjugates of \( f(u) \) and \( \| v \|^2 / (2t) \):

\[ (f_t)^*(y) = f^*(y) + \frac{t}{2} \| y \|^2 \]

- hence, conjugate is strongly convex with parameter \( t \)
Gradient of Moreau envelope

\[ f_t(x) = \sup_y \left( x^T y - f^*(y) - \frac{t}{2} ||y||^2 \right) \]

- maximizer in definition is unique and satisfies

\[ x - ty \in \partial f^*(y) \iff y \in \partial f(x - ty) \]

- maximizing \( y \) is the gradient of \( f_t \): from pages 4.7 and 6.4,

\[ \nabla f_t(x) = \frac{1}{t} \left( x - \text{prox}_{t f}(x) \right) = \text{prox}_{(1/t)f^*}(x/t) \]

- gradient \( \nabla f_t \) is Lipschitz continuous with constant \( 1/t \) (see page 5.19 or 4.9)
Interpretation of proximal point algorithm

apply gradient method to minimize Moreau envelope

\[
\text{minimize } f(t)(x) = \inf_u \left( f(u) + \frac{1}{2t} \|u - x\|_2^2 \right)
\]

this is an **exact** smooth reformulation of problem of minimizing \( f(x) \):

- solution \( x \) is minimizer of \( f \)
- \( f(t) \) is differentiable with Lipschitz continuous gradient (\( L = 1/t \))

**Gradient update:** with fixed \( t_k = 1/L = t \)

\[
x_{k+1} = x_k - t \nabla f(t)(x_k) = \text{prox}_{tf}(x_k)
\]

… the proximal point update with constant step size \( t_k = t \)
Interpretation of augmented Lagrangian algorithm

\[
\text{minimize } f(x) + g(Ax)
\]

- augmented Lagrangian iteration is

\[
(\hat{x}, \hat{y}) = \arg\min_{x,y} \left( f(x) + g(y) + \frac{t}{2} \|Ax - y + (1/t)z\|_2^2 \right)
\]

\[
z := z + t(A\hat{x} - \hat{y})
\]

- with fixed \( t \), dual update is gradient step applied to a smoothed dual

- after eliminating \( y \), primal step can be written as

\[
\hat{x} = \arg\min_x \left( f(x) + g(1/t) (Ax + (1/t)z) \right)
\]

- second term \( g(1/t)(Ax + (1/t)z) \) is a smooth approximation of \( g(Ax) \)

- adding the offset \( z/t \) allows us to use a fixed \( t \)
Example

\[ \text{minimize} \quad f(x) + \|Ax - b\|_1 \]

- augmented Lagrangian iteration is

\[ (\hat{x}, \hat{y}) = \arg\min_{x, y} \left( f(x) + \|y\|_1 + \frac{t}{2} \|Ax - y + (1/t)z\|_2^2 \right) \]

\[ z := z + t(A\hat{x} - \hat{y}) \]

- primal step after eliminating \( y \): \( \hat{x} \) is the solution of

\[ \text{minimize} \quad f(x) + \phi_{1/t} (Ax - b + (1/t)z) \]

with \( \phi_{1/t} \) the Huber penalty applied componentwise (page 8.12)
References

Accelerated proximal point algorithm


Augmented Lagrangian algorithm