17 Quadratic optimization

Thursday, October 29, 2020 7:39 PM

Quadratic Optimization (Ch 6)

Let's consider in detail a couple common classes of quadratic optimization, before moving on to general results in optimization theory.

We will discuss minimizing $Q(x) = \frac{1}{2}x^TAx - x^Tb$ over
(1) $x \in \mathbb{R}^n$
(2) $x \in \mathbb{R}^n$ subject to linear or affine constraints
(3) x in the unit sphere.

This is important in practice because many energy functions can be defined in this form.

Def. 6.2 Given any symmetric $A \in \mathbb{R}^{n\times n}$, we write $A \succeq 0$ if A is pos. semi def.
$A \succ 0$ if A is pos. def.

Also, $A \succeq B$ if $A - B \succ 0$, a partial order on matrices called the pos. semidef. cone ordering.

Prop 6.2 Given a quadratic function,

$$Q(x) = \frac{1}{2} x^TAx - x^Tb,$$

if A is symmetric and $A \succ 0$, then $Q(x)$ has a unique global minimum at the solution of the linear system $Ax = b$. The minimum value of $Q(x)$ is

$$Q(A^{-1}b) = -\frac{1}{2} b^TA^{-1}b.$$

proof. Let $x = A^{-1}b$. Let $y \in \mathbb{R}^n$.

Then
$$\begin{aligned} Q(y) - Q(x) &= \frac{1}{2}y^TAy - y^Tb - \frac{1}{2}x^TAx + x^Tb \\ &= \frac{1}{2}y^TAy - y^TAx + \frac{1}{2}x^TAx \\ &= \frac{1}{2}(y-x)^TA(y-x) \geq 0 \end{aligned}$$

$$\Rightarrow Q(y) \geq Q(x).$$

$$\Rightarrow \min_{x \in \mathbb{R}} Q(x) = Q(A^{-1}b) = -\frac{1}{2}b^TA^{-1}b.$$

$\boxtimes$

Aside: If $Q(x) = \frac{1}{2}x^TAx - x^Tb + c$, then $\underset{x \in \mathbb{R}^n}{\arg\min}\, Q(x) = A^{-1}b$, but $Q(A^{-1}b) = -\frac{1}{2}b^TA^{-1}b + c$.

This allows us to recast a linear problem $Ax = b$ as a variational problem (finding the min. of an energy function). Often, we have additional constraints.

Def. 6.3 The quadratic constrained minimization problem consists in minimizing

$$Q(x) = \frac{1}{2}x^TA^{-1}x - b^Tx \quad \text{subject to linear constraints } B^Tx = f,$$

where $A^{-1} \in \mathbb{R}^{m\times m}$ is SPD, $B \in \mathbb{R}^{m\times n}$ has rank n, and where $b, x \in \mathbb{R}^m$ and $f \in \mathbb{R}^n$.

Note that we use A^{-1} instead of A because this constrained minimization has an interpretation as a set of equilibrium equations that give A. Notation taken from [Strang 1986].

The matrix $K = B^TAB$ is the stiffness matrix of e.g. a spring-mass system, or electrical networks, etc.

Recall that we can use Lagrange multipliers to solve this. The Lagrangian of the system is $L(x,\lambda) = Q(x) + \lambda^T(B^Tx - f) = \frac{1}{2}x^TA^{-1}x - (b - B\lambda)^Tx - \lambda^Tf$.

A necessary condition is $\nabla L(x,\lambda) = 0$.

$$\begin{cases} \dfrac{\partial L}{\partial x}(x,\lambda) = A^{-1}x - (b - B\lambda) = 0 \\[2ex] \dfrac{\partial L}{\partial \lambda}(x,\lambda) = B^Tx - f = 0 \end{cases}$$

$$\Rightarrow \left.\begin{aligned} A^{-1}x + B\lambda &= b \\ B^Tx &= f \end{aligned}\right\} \begin{pmatrix} A^{-1} & B \\ B^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}$$

$$\Rightarrow x = A(b - B\lambda)$$

$$\Rightarrow B^TA(b - B\lambda) = f$$

$$\Rightarrow B^TAB\lambda = B^TAb - f$$

$$\Rightarrow \lambda = (B^TAB)^{-1}(B^TAb - f), \quad x = A(b - B\lambda).$$

(If we let $e = b - B\lambda$, we get equilibrium equations

$$\begin{cases} e = b - B\lambda \\ x = Ae \end{cases}$$

[Strang, 1986])

$$\begin{cases} e = b - B\lambda \\ x = Ae \\ B^T x = f \end{cases}$$ [Strang, 1986]

Let us define the dual function $G(\lambda)$ as follows:

$$G(\lambda) = \frac{1}{2}(B\lambda - b)^T A (B\lambda - b) + \lambda^T f.$$

Note that $\min_x L(x,\lambda) = L(A(b - B\lambda), \lambda)$ (by Prop 6.2)

$$= -G(\lambda).$$

Clearly, $L(x,\lambda) \geq -G(\lambda) \quad \forall x, \lambda$ because we minimized over x to get G.

But when $B^T x = f$, $L(x,\lambda) = Q(x)$, so

$$\forall \lambda, \quad \min_{x | B^T x = f} Q(x) = \min_{x | B^T x = f} L(x,\lambda) \geq \min_x L(x,\lambda) = -G(\lambda)$$

$$\Rightarrow \quad \min_{x | B^T x = f} Q(x) \geq \max_\lambda -G(\lambda).$$

We are seeing here a special case of duality, which we will cover in more detail later.

Prop. 6.3 The quadratic constrained minimization problem has a unique solution (x, λ) given by

$$\begin{pmatrix} A^{-1} & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}$$

Furthermore, the component λ of the above solution is the unique value for which $-G(\lambda)$ is maximum.

proof. Let's compute $Q(x) + G(\lambda)$ subject to $B^T x = f$.

$$Q(x) + G(\lambda) = \frac{1}{2} x^T A^{-1} x - b^T x + \frac{1}{2}(B\lambda - b)^T A (B\lambda - b) + \lambda^T f$$

$$= \frac{1}{2}(A^{-1}x + B\lambda - b)^T A (A^{-1}x + B\lambda - b) \geq 0$$

and $Q(x) + G(\lambda) = 0$ iff $A^{-1}x + B\lambda - b = 0$

$$\Leftrightarrow A^{-1}x + B\lambda = b.$$

Then $Q(x) = -G(\lambda)$ exactly when $A^{-1}x + B\lambda = b$ (i.e. $Q(x) = \min_{x | B^T x = f} Q(x)$)

But $\min_{x | B^T x = f} Q(x) \geq \max_\lambda -G(\lambda)$ from above, so equality is achieved precisely at a constrained minimum of Q and an unconstrained maximum of $-G$. ☑

$x \mid Bx=f$ λ

precisely at a constrained minimum of Q and an unconstrained maximum of $-G$.

Prop. 6.4 If A is an invertible square matrix, then the function

$$f(x)=\frac{1}{2}x^TAx - x^Tb$$

has a minimum value iff $A \succeq 0$, in which case this optimal value is obtained for a unique value of x, namely $x^* = A^{-1}b$, and with

$$f(A^{-1}b) = -\frac{1}{2}b^TA^{-1}b.$$

proof. Note $\frac{1}{2}(x-A^{-1}b)^TA(x-A^{-1}b) = \frac{1}{2}x^TAx - x^Tb + \frac{1}{2}b^TA^{-1}b$.

Thus, $f(x) = -\frac{1}{2}x^TAx - x^Tb = \frac{1}{2}(x-A^{-1}b)^TA(x-A^{-1}b) - \frac{1}{2}b^TA^{-1}b$.

If A has a neg. eigenvalue $-\lambda$ $(\lambda>0)$, with eigenvector u, then $\forall\, \alpha\in\mathbb{R}$, $\alpha\neq 0$, if $x=\alpha u + A^{-1}b$, then since $Au=-\lambda u$,

$$f(x) = \frac{1}{2}\alpha u^TA\alpha u - \frac{1}{2}b^TA^{-1}b$$

$$= -\frac{1}{2}\alpha^2\lambda\|u\|_2^2 - \frac{1}{2}b^TA^{-1}b.$$

Thus, f has no minimum as we can set $\alpha\to\infty$, sending $f(x)\to-\infty$.

Thus, in order to have a minimum, $A\succeq 0$.

But A is invertible, so $A\succ 0$.

$$\Rightarrow (x-A^{-1}b)^TA(x-A^{-1}b) > 0 \text{ iff } x\neq A^{-1}b, \text{ and}$$

$$= 0 \text{ iff } x = A^{-1}b, \text{ which is the minimum of } f.$$

Prop. 6.5 If A is a symmetric $n\times n$ matrix, then the function

$$f(x) = \frac{1}{2}x^TAx - x^Tb$$

has a minimum iff $A\succeq 0$ and $(I-AA^+)=0$, in which case this min is

$$p^* = -\frac{1}{2}b^TA^+b.$$

Furthermore, if A is diagonalized as $A=U^T\Sigma U$ (U orthogonal), then the optimal value is achieved by all $x\in\mathbb{R}^n$ of the form

$$x = A^+b + U^T\begin{pmatrix}0\\z\end{pmatrix}, \quad \text{for any } z\in\mathbb{R}^{n-r},\ r=\text{rank}(A).$$

proof idea: Relatively long, but makes use of things like SVD and projections.

This is all in the unconstrained setting. Can we say anything given constraints?

Aside: Consider $f(x) = \frac{1}{2} x^T A x - x^T b$ subject to linear constraints $C^T x = 0$

(or affine constraints $C^T x = t$) $C \in \mathbb{R}^{n \times m}$.

First permute C by taking $C\Pi^{-1}$, where Π^{-1} is a permutation matrix s.t. the first $r = \text{rank}(C)$ cols of $C\Pi^{-1}$ are linearly ind.

Then $C\Pi^{-1} = Q^T \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix}$, a QR-decomposition, where $Q^T \in \mathbb{R}^{n \times n}$ is orthogonal, and $R \in \mathbb{R}^{r \times r}$ invertible upper triangular, $S \in \mathbb{R}^{r \times (m-r)}$.

$$\Rightarrow \quad C = Q^T \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix} \Pi.$$

Then let $x = Q^T \begin{pmatrix} y \\ z \end{pmatrix}$, $y \in \mathbb{R}^r$, $z \in \mathbb{R}^{n-r}$.

$$\Rightarrow C^T x = \Pi^T \begin{pmatrix} R^T & 0 \\ S^T & 0 \end{pmatrix} Q x = \Pi^T \begin{pmatrix} R^T & 0 \\ S^T & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = 0$$

$\Rightarrow y = 0$ and every sol of $C^T x$ has the form

$$x = Q^T \begin{pmatrix} 0 \\ z \end{pmatrix}.$$

Then $\min f(x) = \frac{1}{2} x^T A x - x^T b$ subject to $C^T x = 0$

becomes minimize $\frac{1}{2} (y^T \; z^T) Q A Q^T \begin{pmatrix} y \\ z \end{pmatrix} + (y^T \; z^T) Q b$

subject to $y = 0$, $y \in \mathbb{R}^r$, $z \in \mathbb{R}^{n-r}$.

$y = 0$ is clearly a much simpler constraint than $C^T x = 0$.

Let $Q A Q^T = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$, $G_{11} \in \mathbb{R}^{r \times r}$, $G_{22} \in \mathbb{R}^{(n-r) \times (n-r)}$

and $Q b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, $b_1 \in \mathbb{R}^r$, $b_2 \in \mathbb{R}^{n-r}$.

letting us rewrite as: minimize $\frac{1}{2} z^T G_{22} z + z^T b_2$, $z \in \mathbb{R}^{n-r}$,

which is an unconstrained optimization.

We can do something similar in the affine case $C^T x = t$.