# How to Find Majorizing Functions? Tricks and an Example

from

Hunter, D. & Lange, K. (2012). A Tutorial on MM Algorithms

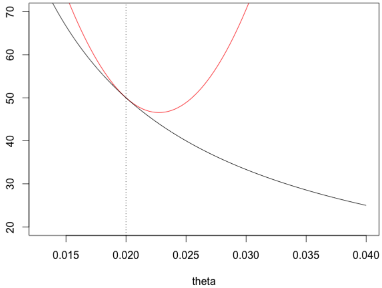
# Recall...

Goal: minimize $f(\theta) : \mathbb{R}^d \mapsto \mathbb{R}$ wrt $\theta \in \mathbb{R}^d$.

Use a majorizing function $g(\theta|\theta^{(m)}) : \mathbb{R}^d \mapsto \mathbb{R}$ such that:

i. (*dominating*) $\quad g(\theta|\theta^{(m)}) \geq f(\theta) \quad \forall\theta$,

ii. (*tangent at* $\theta^{(m)}$) $\quad g(\theta^{(m)}|\theta^{(m)}) = f(\theta^{(m)})$.




black: $f(\theta) = 1/\theta$; red: majorizing function at $\theta^{(m)} = 0.02$

# How to find a majorizing/minorizing function?

# How to find a majorizing/minorizing function?

3.1 Jensen's inequality - gives us the EM algorithm

3.2 Minorization via Supporting Hyperplanes

3.3 Majorization via the Definition of Convexity

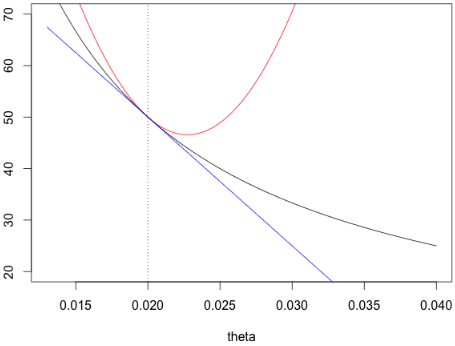
3.4 Majorization via a Quadratic Upper Bound

3.5 The Arithmetic-Geometric Mean Inequality

3.6 The Cauchy-Schwartz Inequality

# How to find a majorizing/minorizing function?

# EX: $f(\theta) = 1/\theta$




red: quadratic upper bound at $\theta^{(m)} = 0.02$
blue: supporting hyperplane (straight line) $\theta^{(m)} = 0.02$

# How to find a majorizing/minorizing function?

# Majorization via the Definition of Convexity

A function $\kappa(t)$ is convex if and only if

$$\kappa\left(\sum_i \alpha_i t_i\right) \leq \sum_i \alpha_i \kappa(t_i),$$

where $\alpha_i \geq 0$ and $\sum \alpha_i = 1$.
Construct $\alpha_i t_i$, for example:

$$\begin{aligned}
f(x'\theta) &= f\left(\sum x_i \theta_i\right) \\
&= f\left(\sum \alpha_i \cdot \left(\frac{x_i(\theta_i - \theta_i^{(m)})}{\alpha_i} + x'\theta^{(m)}\right)\right) \\
&\leq \sum \alpha_i \cdot f\left(\frac{x_i(\theta_i - \theta_i^{(m)})}{\alpha_i} + x'\theta^{(m)}\right)
\end{aligned}$$

# How to find a majorizing/minorizing function?

# How to find a majorizing/minorizing function?

# The Arithmetic-Geometric Mean Inequality

Start with something simple:

$$2x_1x_2 \leq x_1^2 + x_2^2 \quad \Rightarrow \quad x_1x_2 \leq \frac{x_1^2}{2} + \frac{x_2^2}{2}$$

For $x_1, x_2 > 0$, we get:

$$\begin{aligned} x_1x_2 &= x_1\sqrt{\frac{x_2^{(m)}}{x_1^{(m)}}} \cdot x_2\sqrt{\frac{x_1^{(m)}}{x_2^{(m)}}} \\ &\leq x_1^2 \cdot \frac{x_2^{(m)}}{x_1^{(m)}} \cdot \frac{1}{2} + x_2^2 \cdot \frac{x_1^{(m)}}{x_2^{(m)}} \cdot \frac{1}{2} \end{aligned}$$

# NBA example

2002-2003 season, 29 teams

each team $i$: $o_i$ (offensive strength) & $d_i$ (defensive strength)

$o_i + d_i$ measures the overall strength of the team

# NBA example: Model

Observe: for each game, $p_{ij}$ = points scored by team $i$ against team $j$

Assume:

1. different games are independent of each other
2. each team's total point in one game is independent of its opponent's point total

Model:

$$p_{ij} \sim \textit{(indep)} \ \text{Poisson}\left(t_{ij} \cdot e^{(o_i - d_j)}\right)$$

where

$$\begin{aligned} p_{ij} &= \text{points scored by team } i \text{ against team } j, \\ t_{ij} &= \text{time (in minutes) of the game}, \\ o_i &= \text{offensive strength of team } i, \\ d_j &= \text{defensive strength of team } j. \end{aligned}$$

# NBA example: MLE

Find MLE for $\theta = (o_i, d_i)_{i=1,\ldots,29}$ by maximizing the log likelihood

$$\textit{maximize} \quad l(\theta) = \sum_{i,j} \left\{ p_{ij} \cdot (o_i - d_j) - t_{ij} \cdot e^{(o_i - d_j)} + p_{ij} ln(t_{ij}) - ln(p_{ij}!) \right\}$$

$$\Rightarrow \textit{minimize} \quad f(\theta) = \sum_{i,j} \left\{ -p_{ij} \cdot (o_i - d_j) + t_{ij} \cdot e^{(o_i - d_j)} \right\}$$

$$\Rightarrow \textit{majorized by} \quad g(\theta|\theta^{(m)}) = \sum_{i,j} \left\{ -p_{ij} \cdot (o_i - d_j) + t_{ij} \cdot h(\theta|\theta^{(m)}) \right\}$$

$$\textit{where} \quad h(\theta|\theta^{(m)}) = \frac{e^{2o_i}}{e^{o_i^{(m)} + d_j^{(m)}}} \cdot \frac{1}{2} + \frac{e^{o_i^{(m)} + d_j^{(m)}}}{e^{2d_j}} \cdot \frac{1}{2}$$

# NBA example: majorizing function

Want to show: $e^{(o_i - d_j)} \leq h(\theta|\theta^{(m)}) = \frac{e^{2o_i}}{e^{o_i^{(m)} + d_j^{(m)}}} \cdot \frac{1}{2} + \frac{e^{o_i^{(m)} + d_j^{(m)}}}{e^{2d_j}} \cdot \frac{1}{2}$

$$e^{(o_i - d_j)} = e^{o_i} \cdot \frac{1}{e^{d_j}}$$

$$Let \;\; x_1 = e^{o_i}$$

$$x_2 = \frac{1}{e^{d_j}}$$

$$Plug\ in \;\; x_1 x_2 \leq x_1^2 \cdot \frac{x_2^{(m)}}{x_1^{(m)}} \cdot \frac{1}{2} + x_2^2 \cdot \frac{x_1^{(m)}}{x_2^{(m)}} \cdot \frac{1}{2}$$

# NBA example: majorizing function

The majorizing function

$$g(\theta|\theta^{(m)}) = \sum_{i,j} \left\{ -p_{ij} \cdot (o_i - d_j) + \frac{t_{ij}}{2} \cdot \frac{e^{2o_i}}{e^{o_i^{(m)} + d_j^{(m)}}} + \frac{t_{ij}}{2} \cdot \frac{e^{o_i^{(m)} + d_j^{(m)}}}{e^{2d_j}} \right\}$$

is easy to minimize piecewise for each $o_i$ and $d_j$.

# NBA example: in case you are interested...

| Team | $\hat{o}_i + \hat{d}_i$ | Wins | Team | $\hat{o}_i + \hat{d}_i$ | Wins |
|---|---|---|---|---|---|
| Cleveland | −0.0994 | 17 | Phoenix | 0.0166 | 44 |
| Denver | −0.0845 | 17 | New Orleans | 0.0169 | 47 |
| Toronto | −0.0647 | 24 | Philadelphia | 0.0187 | 48 |
| Miami | −0.0581 | 25 | Houston | 0.0205 | 43 |
| Chicago | −0.0544 | 30 | Minnesota | 0.0259 | 51 |
| Atlanta | −0.0402 | 35 | LA Lakers | 0.0277 | 50 |
| LA Clippers | −0.0355 | 27 | Indiana | 0.0296 | 48 |
| Memphis | −0.0255 | 28 | Utah | 0.0299 | 47 |
| New York | −0.0164 | 37 | Portland | 0.0320 | 50 |
| Washington | −0.0153 | 37 | Detroit | 0.0336 | 50 |
| Boston | −0.0077 | 44 | New Jersey | 0.0481 | 49 |
| Golden State | −0.0051 | 38 | San Antonio | 0.0611 | 60 |
| Orlando | −0.0039 | 42 | Sacramento | 0.0686 | 59 |
| Milwaukee | −0.0027 | 42 | Dallas | 0.0804 | 60 |
| Seattle | 0.0039 | 40 | | | |

Thank you!