凸函数 Convex Function

$f: R^n \to R$ 被称为凸函数，当且仅当对于 $\forall x_1, x_2 \in R^n$ 和 $\forall \theta \in [0,1]$，满足下式：

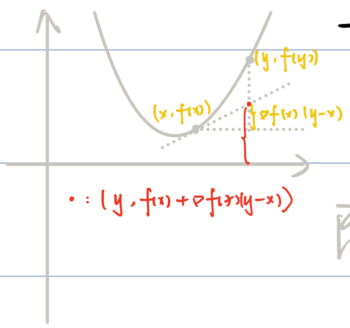
$$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$$

这表示在两点 x_1、x_2 之间的任意点 $\theta x_1 + (1-\theta)x_2$ 上，函数值不会超过这两点连成的直线上的值。

与定义等价

(其定义域一定是开区间)

一阶条件：对于可微函数 f，如果它是凸函数，则其在任意点的梯度 $\nabla f(x)$ 定义了一个支撑超平面，并满足：



$$f(y) \ge f(x) + \nabla f(x)^T (y-x), \quad \forall y \in R^n$$

即 x 处的切线永远在函数下方或与之相切。

与定义等价

△ **二阶条件**：对二阶可导函数，若函数的Hessian矩阵(即二阶导数矩阵)在定义域内的所有点都半正定(即特征值都非负)，则 f 为凸函数。

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in R^n.$$

注意：$\nabla^2 f(x) \succ 0 \Rightarrow f$ 严格凸
反之则不一定(如 $f = x^4, x \in R$)

NOTE：若 dom f 为凸集，且满足上述定义 $\Longleftrightarrow$ f 为凸函数.

① 严格凸，定义中 $f(\cdots) < \cdots$

② 凹：所有条件取反　③ 仿射函数：既凹又凸.

凸函数的扩展

$f: R^n \to R$ 为凸　　$\mathrm{dom}\, f = C \subseteq R^n$

↓ 扩展

$$\tilde{f} = \begin{cases} f(x) & x \in \mathrm{dom}\, f \\ \infty & x \notin \mathrm{dom}\, f \end{cases}$$

显然：$\tilde{f}: R^n \to R$，$\mathrm{dom}\, \tilde{f} = R^n$

① $\tilde{f}$ 的定义域显然是凸集

② 取 $\forall x_1, x_2 \in \mathrm{dom}\, \tilde{f}$

1) x_1, x_2 都在 $\mathrm{dom}\, f$ 内，显然满足定义中不等式.

2) x_1, x_2 一个在 $\mathrm{dom}\, f$，一个不在，其连线上的值为 ∞，满

3) x_1, x_2 都不在 $\mathrm{dom}\, f$，满足

⇓ $\tilde{f}$ 仍为凸函数.

示性函数

设有凸集 $C \subseteq R^n$，其示性函数定义为：

① $$f_C(x) = \begin{cases} 无定义 & x \notin C \\ 0 & x \in C \end{cases}$$

① f_C 定义域，凸.（$\because \mathrm{dom}\, f_C = C$）

② 不等式显然成立　$\Rightarrow$　f_C 凸.

⇓ 扩展

② $I_C(x) = \begin{cases} \infty & x \notin C \\ 0 & x \in C \end{cases}$

显然 I_C 也是凸函数.

另:

③ $J_C(x) = \begin{cases} 1 & x \notin C \\ 0 & x \in C \end{cases}$

J_C 既非凸也非凹函数. 因此, 扩展时函数值一定要设为 ∞.

一阶条件的证明:

① n 取 1 时, 考虑一维情况, 对于可微的 f

$f: R \to R$ 为凸 $\Longleftrightarrow$ dom f 为凸集, 且

$$f(y) \geq f(x) + f'(x)(y-x)$$

1) $\Rightarrow$ 方向证明: f 为凸, 则 dom f 为凸, 显然.

对 $\forall x, y \in$ dom f

该凸组合不涉及 α 有 $t \in [0,1]$, 使得 $x+t(y-x) \in$ dom f

由凸函数定义, 下式成立.

$$f(x+t(y-x)) \leq (1-t)f(x) + tf(y)$$

$$\therefore tf(y) \geq tf(x) + f(x+t(y-x)) - f(x)$$

对 t 除, $\therefore$ $f(y) \geq f(x) + \dfrac{f(x+t(y-x)) - f(x)}{t}$ 这里 $t(y-x)$ 为极小值,

两端取极限令 $\lim\limits_{t\to 0^+}$，$f(y) \geq f(x) + \lim\limits_{t\to 0} \dfrac{f(x)+f'(x)\cdot t(y-x)+o(t) - f(x)}{t}$ 做泰勒展开.

即 $f(y) \geq f(x) + f'(x)(y-x)$

2) $\Leftarrow$ 方向证明，设 $\forall x, y \in \mathrm{dom} f$ 且 $x \neq y$

取 $\theta \in [0,1]$，构造凸组合 $z = \theta x + (1-\theta) y \in \mathrm{dom} f$

已知 $f(x) \geq f(z) + f'(z)(x-z)$ $\quad \cdot\ \theta$

$f(y) \geq f(z) + f'(z)(y-z)$ $\quad \cdot\ (1-\theta)$

上面两不等式分别两端同乘相应系数后相加.

$\theta f(x) + (1-\theta) f(y) \geq f(z) + [\theta x + (1-\theta) y - z] f'(z)$

又由 $z = \theta x + (1-\theta) y$，则下式成立：

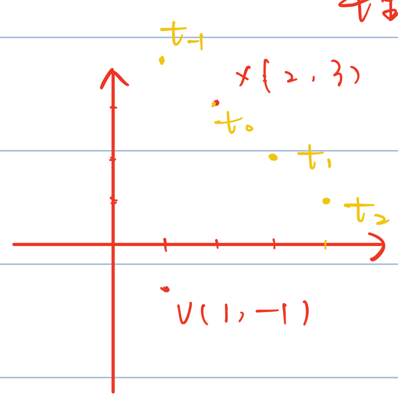
$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$

凸函数定义2： $f: R^n \to R$ 为凸 $\Leftrightarrow$ ① $\mathrm{dom} f$ 凸集

② 对 $\forall x \in \mathrm{dom} f$，$\forall v$ 找一个任意方向 v

都满足 $g(t) = f(x+tv)$ 为凸，$\mathrm{dom}\, g = \{t \mid x+tv \in \mathrm{dom} f\}$

$t \in R$



理解：给定 x、$v \in R^2$，从 t 的变化过程来观察，可以从 $t_{-1} \to t_2$ 的变化看 $x+tv$ 的轨迹. 对于 $f(x+tv)$ 而言，沿这条轨迹对"铁锅"做切割（想象 f 是一个三维凸形状），可得一条切线，若对 $\forall x \in \mathrm{dom} f$ 和 $\forall v$ 而言，这条切

线都是凸的，那么f就是凸的。

利用定义2证明一阶条件高维的情况

$f: R^n \to R, \quad n > 1$

① 若f凸 $\Rightarrow$ $\forall x, y \in \mathrm{dom} f$，有 $f(y) \geq f(x) + \nabla f^T(x)(y-x)$

f为凸，$\forall x, y \in \mathrm{dom} f$

构造 $g(t) = f(ty + (1-t)x)$　　对x、y仿射组合.

$= f(x + t(y-x))$　　方向v即 $y-x$

有 $g'(t) = \nabla f^T(ty + (1-t)x)(y-x)$

由f凸可得g(t)凸：$\forall t_1, t_2 \in \mathrm{dom}\, g$

有 $g(t_1) \geq g(t_2) + g'(t_2)(t_1 - t_2)$　　低维情况已证明.

$g(1) \geq g(0) + g'(0)$ 显然成立.

$\therefore f(y) \geq f(x) + \nabla f^T(x)(y-x)$

② 若 $\forall x, y \in \mathrm{dom} f$，$f(y) \geq f(x) + \nabla f^T(x)(y-x) \Rightarrow$ f凸.

选择 t、$\tilde{t}$ 使得 $ty + (1-t)x \in \mathrm{dom} f$

$\tilde{t}y + (1-\tilde{t})x \in \mathrm{dom} f$

有 $f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x) + \nabla f^T(\tilde{t}y + (1-\tilde{t})x)(y-x)(t-\tilde{t})$

构造 $g(t) = f(ty + (1-t)x)$，$g(\tilde{t}) = f(\tilde{t}y + (1-\tilde{t})x)$

有 $g'(\tilde{t}) = \nabla f^T(\tilde{t}y + (1-\tilde{t})x)(y-x)$

$\therefore g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t - \tilde{t})$　　低维情况已证明.

∴ g 凸. 又由 x、y、t 的任意性　　由定义2

∴ f 凸.

重要凸函数 / 凸函数讨论.

一 ① 定义二次函数，$f: R^n \to R$，$\mathrm{dom} f = R^n$

$$f(x) = \frac{1}{2}x^T P x + q^T x + r$$

行向量

$P \in S^n$ (对称矩阵)　$q^n \in R^n$，$r \in R$

$\Rightarrow \nabla^2 f(x) = P$，即若 $P \in S^n_+ \Rightarrow f$ 一般凸

$P \in S^n_{++} \Leftrightarrow f$ 严格凸.

$P \in S^n_- \Rightarrow f$ 凹.

考虑：若 f 严格凸，那 P 一定正定吗？ 半正定 ✓

× ② $f(x) = \frac{1}{x^2}$，$x \neq 0$ 且 $x \in R$

易得 $f''(x) = 6x^{-4} > 0$，那么 f(x) 应为凸，对吗？

不对，一定要注意 dom f 必须为凸集，否则不是凸函数.

√ ③ 仿射函数　$f(x) = Ax + b$

$\nabla f^2(x) = 0$　零矩阵

∴ 即凸又凹.

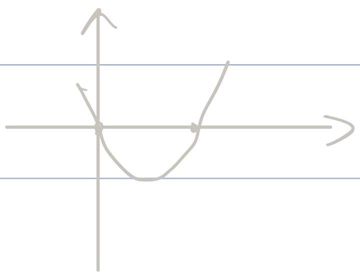
√④ 指数函数 $f(x)=e^{ax}$, $x \in \mathbb{R}$, a 为任意实数

$\nabla f^2(x) = a^2 e^{ax} \geq 0$

∴ 凸函数.

— ⑤ 幂函数 ① $f(x) = x^a$, $x \in \mathbb{R}_{++}$

$\nabla f^2(x) = a(a-1)x^{a-2}$



1) $a \in [0,1]$ 时, $\nabla f^2(x) \leq 0$

2) $a \geq 1$ 或 $a \leq 0$, $\nabla f^2(x) \geq 0$

∴ $a=1$ 时, 既凹又凸.

$a=0$ 时, 同上.

② $f(x) = |x|^p$, $x \in \mathbb{R}$

$$f'(x) = \begin{cases} p x^{p-1} & x \geq 0 \\ -p(-x)^{p-1} & x < 0 \end{cases} \quad , \; p \neq 0, -1$$

$p=1$ 时 $f(x)$ 不可导.
但 f 显然是凸函数

$$f''(x) = \begin{cases} p(p-1)x^{p-2} & x \geq 0 \\ p(p-1)x^{p-2} & x < 0. \end{cases}$$

注意. $p \geq 2$ 时才二阶可微.

1) 当 $p \geq 1$ 时, f 凸.

2) 当 $p < 1$ 时, 没有一定的结论, 需具体讨论.

一 ⑥ 对数函数 $f(x) = \log_a(x)$ ， $x \in R_{++}$ ， $a>0$ 且 $a \neq 1$

$f'(x) = \frac{1}{x \ln a}$

$f''(x) = -\frac{1}{x^2 \ln a}$

∴ 1) $a \in (0,1)$ 时， f 凸

2) $a > 1$ 时， f 凹.

✓ ⑦ 负熵 $f(x) = x \underline{\log x}$ ， $x \in R_{++}$

$\ln x$

$f'(x) = \underline{\log x} + 1$

$\ln x$

$f''(x) = \frac{1}{x} > 0$

∴ f 为严格凸. (将 $\ln x$ 乘上 x 后，变成了凸函数)

✓ ⑧ 范数 R^n 空间的范数 $P(x)$ ， $x \in R^n$

满足三性质：

1) $P(ax) = |a| P(x)$ 缩放性质

2) $P(x+y) \leq P(x) + P(y)$

3) $P(x) = 0 \iff x = 0$

利用定义证明其凸性：

$\forall x, y \in R^n$ ， $\forall \theta \in [0,1]$

$P(\theta x + (1-\theta)y) \leq P(\theta x) + P((1-\theta)y)$

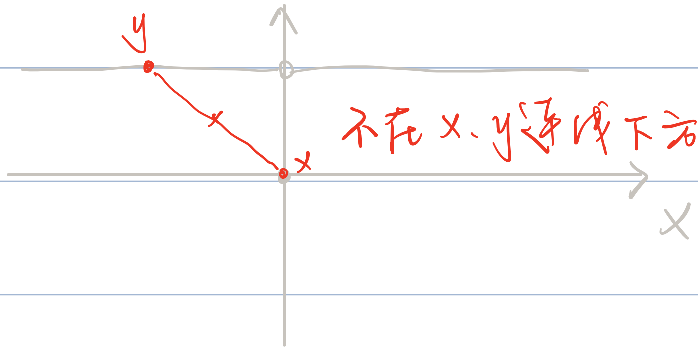
∴ $P(\theta x + (1-\theta)y) \leq \theta P(x) + (1-\theta) P(y)$

∴ 范数一定为凸函数.

× ⑨ 零范数 $\|x\|_0$ = 非零元素的数目.

既非范数，也非凸函数.

1) $x\in R$ 时



不在 x、y 连线下方 => 非凸.

2) 取 $a=2,\ x=1$

$\|ax\|_0 = \|2\|_0 = 1$ ， $|a|\cdot\|x\|_0 = 2\|1\|_0 = 2$

二者不相等 => $\|x\|_0$ 也不是范数.

✓ ⑩ 极大值函数 $f(x) = \max\{x_1, \cdots, x_n\},\ x\in R$

$\forall x, y \in R^n,\ \forall\theta\in[0,1]$

$f(\theta x+(1-\theta)y) = \max\{\theta x_i + (1-\theta)y_i,\ i=1\cdots n\}$

$\le \theta\max\{x_i, i=1\cdots n\} + (1-\theta)\max\{y_i, i=1\cdots n\}$

$\le \theta f(x) + (1-\theta) f(y)$

∴ f 凸.

解析逼近: log-sum-up: $g(x) = \log(e^{x_1}+\cdots+e^{x_n}),\ x\in R^n$

显然 g 二阶可导。 $\max\{x_1\cdots x_n\} \le g(x) \le \max\{x_1\cdots x_n\} + \underbrace{\log n}_{\text{近似误差}}$

且 g 为凸函数 $\frac{\partial g}{\partial x_i} = \frac{e^{x_i}}{e^{x_1}+\cdots+e^{x_n}} \xRightarrow{\text{Hessian}} H = \begin{bmatrix} H_{11} & \cdots & \cdots \\ \cdots & H_{ij} & \cdots \\ \cdots & \cdots & H_{nn} \end{bmatrix}$ $H_{ij} = \begin{cases} \cdots & i\ne j \\ \cdots & i=j \end{cases}$

设 $z = [e^{x_1}, \cdots, e^{x_n}]^T$ （列向量）

可得 H 的简化形式:

$$H = \frac{1}{(1^T z)^2}\left[\underbrace{(1^T z)}_{\text{标量}} \underbrace{\mathrm{diag}\{z\}}_{\text{展开为对角矩阵}} - z z^T\right]$$

记为 matrix K, $K \in R^{n\times n}$

NOTE: ? $\Rightarrow$ 半正定矩阵

对 $\forall v \in R^n$, $v^T M v \geq 0 \Rightarrow M$ 半正定

有: $v^T K v = (1^T z)\, v^T \mathrm{diag}\{z\} v - v^T z z^T v$

$$= \left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2$$

定义 $a_i = v_i\sqrt{z_i}$, $b_i = \sqrt{z_i}$

$$= (b^T b)(a^T a) - (a^T b)^2 \geq 0$$

NOTE: Cachy-schwartz 不等式

(两个向量的内积绝对不超过他们范数的乘积)

$\therefore H$ 半正定 $\Rightarrow$ log-sum-up 为凸函数

⑪ $\min_x \max_y f(x, y)$, 若 f 性质较好 $\Rightarrow$ 凸.

⑫ 几何平均

$$f(x) = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{\frac{1}{n}}, \quad x \in R^n_{++}$$

证明 $f(x)$ 为凹

可以考虑先对等式两端取对数

$g(x) = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i)$ 由于g单调，若g凹 => f凹.

对g(x)求Hessian矩阵，判断其正定性可知

g凹. => f凹.

⑬ 行列式的对数

$f(x) = \log \underline{\det(X)}$ 取行列式 $\quad \mathrm{dom} f = S_{++}^n$ 保证 $\det(X) > 0$

当 $n=1$ 时，f显然凹.

$n>1$ 时：$\underbrace{\forall Z \in S_{++}^n}_{\text{起点}}$，$\underbrace{\forall t \in R}_{\text{新自变量}}$，$\underbrace{\forall V \in R^{n\times n}}_{\text{方向}}$

且满足 $Z + tV \in S_{++}^n = \mathrm{dom} f$

∴要求 $V \in S^n$ （n阶对称矩阵）

$$\therefore g(t) = f(Z+tV) = \log\det(Z+tV)$$

$$= \log\det\left\{Z^{\frac{1}{2}}\left(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}\right)Z^{\frac{1}{2}}\right\}$$

$$= 2\log\det(Z^{\frac{1}{2}}) + \log\det\left(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}\right)$$

$$= \log\det(Z) + \sum_{i=1}^{n}\log(1+t\lambda_i) \quad ?$$

?: λ_i : $tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}$ 的特征值 （第i个）

由于 $Z, V \in S^n$，有 $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}} \in S^n$

对其做分解 $\Rightarrow Q\Lambda Q^T$ 且 $QQ^T = I$

$$\therefore \det(I + tZ^{-\frac{1}{2}} V Z^{-\frac{1}{2}})$$
$$= \det(QQ^T + tQ\Lambda Q^T)$$
$$= \underline{\det(Q)}\det(I+t\Lambda)\underline{\det(Q^T)}$$
$$= \det(I+t\Lambda)$$
$$= \prod_{i=1}^{n}(1+t\lambda_i)$$ 连乘.

讨论 $g(t)$ 的凹凸:

$$g'(t) = \sum_i \frac{\lambda_i}{1+t\lambda_i}$$

$$g''(t) = \sum_i \frac{-\lambda_i^2}{(1+t\lambda_i)^2} \leq 0$$

$\therefore g$ 凹 $\Rightarrow f$ 凹

保持函数的凸性.

非负加权和

① $f_1 \cdots f_m$ 为凸，则 $f = \sum_{i=1}^{m} w_i f_i$ 为凸 (若 $w_i \geq 0$, $\forall i$)

① 定义域的凸性:

f_i 的定义域的交集 $\Rightarrow$ 凸集

② f 满足凸函数的定义: 证明略.

这里相当于指定了 y, 变量仅为 x.

② 若 $f(x, y)$, 对 $\forall y \in A$, $\underline{f(x,y)}$ 均为凸.

NOTE：f不要求为凸函数，A为某个集合.

若f对(x,y)这个向量是jontly convex的，则

②成立.

设 $w(y) \geq 0$，$\forall y \in A$：

那么 $g(x) = \int_{y\in A} w(y) f(x,y)\, dy$ 为凸.

仿射映射

① $f: R^n \to R$，$A \in R^{n\times m}$，$b \in R^n$

$g(x) = f(Ax+b)$，$x \in R^m$，$\mathrm{dom}\, g = \{x \mid Ax+b \in \mathrm{dom}\, f\}$

又先仿射，再给f作用

若f凸，则g凸.

证：令 $x, y \in \mathrm{dom}\, g$，$\forall \theta \in [0,1]$

$$
\begin{aligned}
g(\theta x+(1-\theta)y) &= f(A(\theta x+(1-\theta)y)+b) \\
&= f(\theta Ax+(1-\theta)Ay+b) \\
&= f(\theta(Ax+b)+(1-\theta)(Ay+b)) \\
&\leq \theta f(Ax+b)+(1-\theta)f(Ay+b) \\
&= \theta g(x)+(1-\theta)g(y) \Rightarrow g\text{凸.}
\end{aligned}
$$

② $f_i: R^n \to R$，$i=1 \cdots m$，对 $\forall i$，f_i 凸.

令 $A \in R^n$，$b \in R$，$g(x) = A^T[f_1(x), \cdots, f_m(x)]^T + b$

先给f变换，再仿射.　　行向量　　列向量

$g(x)$ 不一定仍是凸，因为这个加权和中，A的每个分量不一定都非负。

两个函数的极大值函数

① f_1, f_2 为凸，定义 $f(x) = \max\{f_1(x), f_2(x)\}$

$$\mathrm{dom}\, f = \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2$$

则 $f(x)$ 一定为凸函数（定义域 $\mathrm{dom}\, f$ 一定是凸集）

证：$\forall x, y \in \mathrm{dom}\, f$，$\forall \theta \in [0,1]$

$$\begin{aligned}
f(\theta x + (1-\theta)y) &= \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\
&\le \max\{\theta f_1(x) + (1-\theta) f_1(y), \theta f_2(x) + (1-\theta) f_2(y)\} \\
&\le \max\{\theta f_1(x), \theta f_2(x)\} + \max\{(1-\theta) f_1(y), (1-\theta) f_2(y)\} \\
&= \theta f(x) + (1-\theta) f(y)
\end{aligned}$$

② 将 max 推广到多个凸函数的 max，仍然成立。

例：向量中 r 个最大元素的和，$x \in R^n$

记 $x[i]$ 为第 i 大元素，

则有 $x[1] \ge \cdots \ge x[r] \ge \cdots$

记和为 $f(x) = \sum_{i=1}^{r} x[i]$

证明其凸性：$f(x)$ 可以写为如下形式

$f(x) = \max\{\underbrace{x_{i_1} + \cdots + x_{i_r}} \mid 1 \le i_1 \cdots \le i_r \le n\}$

从 $x \in R^n$ 中取出r个元素并求和，共 C_n^r 种取法

可看作对x的仿射映射，共 C_n^r 种映射，

每一种映射可看作是 $f_i(x)$，对其取max，仍凸.

无限个凸函数的极大值：

$f(x,y)$ 对于x为凸，$\forall y \in A$（指定y）

$g = \sup_{y \in A} f(x,y)$ 也是凸函数.

取极大（supreme）

例：实对称矩阵的最大特征值

$f(x) = \lambda_{max}(X)$，$\mathrm{dom}\, f = S^n$

特征值满足：$Xy = \lambda y$ 特征向量

$y^T X y = y^T \lambda y$

$y^T X y = \lambda \|y\|_2^2$

$\therefore \lambda = \dfrac{y^T X y}{\|y\|_2^2}$ （有特征向量即可找到对应特征值）.

进一步，将y单位化，则有：

$\lambda_{max}(X) = \sup\{y^T X y \mid \|y\|_2 = 1\}$

$\because X \in S^n$（全体 $n \times n$ 方阵，为凸集） 对每一个规范化的 y_i，设想 $f_i(x) = y_i^T X y_i$，取max仍凸.

$\therefore$ 对于 $\forall y$，$y^T X y$ 都是对于X的凸函数.

$\therefore f(x)$ 为凸函数.

$g(x)$

验证 $\forall\, x, y, \theta \in [0,1]$

$g(\theta x + (1-\theta) y)$

$= a^T (\theta X + (1-\theta) y) a$

$= \theta a^T X a + (1-\theta) a^T y a$

$= \theta g(X) + (1-\theta) g(y)$