Euristica

Vogliamo sfruttare il fatto che la propagazione del calore effettua un rilassamento; in particolare ci rifacciamo al caso del problema di cauchy su $\mathbb{R}$

Sfruttando l'azione rilassante della propagazione del calore, cerchiamo di utilizzare l'idea che l'informazione relativa ai punti di ottimo viaggi in maniera tanto piu predominante tanto maggiore é la magnitudine del punto, sovrastando quindi l'informazione portata dai punti di ottimo locali ma meno desiderabili.

Idea di dimostrazione ; provare a dimostrare che i punti di massimo locale hanno derivata rispetto al tempo negativa e i punti di minimo locale hanno derivata rispetto al tempo negativa e che la magnitudine di questa derivata dipende in una qualche misura dall'entitá del massimo. Dal momento che la derivata misura , in un certo senso , la velocitá di decrescita/crescita di un massimo/minimo , provare quindi che gli ultimi a "sopravvivere" sono i massimi / minimi globali

$$\partial_t \left\{ \frac{1}{\sqrt{t}} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) \right\}$$
$$\partial_t \left\{ \frac{1}{\sqrt{t}} \right\} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) + \frac{1}{\sqrt{t}} \partial_t \left\{ \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) \right\}$$
$$\partial_t \left\{ \frac{1}{\sqrt{t}} \right\} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) + \frac{1}{\sqrt{t}} \int_{\mathbb{R}} f(y) \partial_t exp\left(-\frac{(y-x)^2}{t}\right)$$
$$-\frac{1}{2t^{\frac{3}{2}}} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) + \frac{1}{\sqrt{t}} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) \frac{1}{t^2}$$

Rilassamento della funzione di costo

Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione di cui cerchiamo un punto di ottimo globale x^*
Definiamo la funzione rilassata come

$f_t(x) := \int_{\mathbb{R}} f(y) \Phi_k(y-x,t) dy$
Con $\Phi_k(z,t) := \frac{1}{\sqrt{4\pi kt}} exp\left(-\frac{z^2}{4kt}\right)$

Poniamo per semplicitá $\hat{k} = \frac{1}{4\pi}$ e otteniamo
$\Phi(z,t) := \Phi_{\hat{k}}(z,t) = \frac{1}{\sqrt{t}} exp\left(-\frac{z^2}{t}\right)$

Sostituendo in $f_t(x)$ otteniamo
$$f_t(x) = \int_{\mathbb{R}} f(y) \Phi(y-x,t) dy$$
$$= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{t}} exp\left(-\frac{(y-x)^2}{t}\right)$$
$$= \frac{1}{\sqrt{t}} \int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right)$$

Convoluzione stocastica

La densitá della distribuzione gaussiana $\mathcal{N}(\mu,\sigma^2)$ puo essere modellata come
$$p_{\mathcal{N}(\mu,\sigma)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\left(\frac{y-\mu}{\sqrt{2}\sigma}\right)^2\right\}$$
$$= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2\sigma^2}} exp\left\{-\frac{(y-\mu)^2}{(\sqrt{2}\sigma)^2}\right\}$$

Se nomino
$t := 2\sigma^2 \implies \sigma^2 = \frac{1}{2}t$
$x := \mu \implies \mu = x$

Otteniamo
$$p_{\mathcal{N}(\mu,\sigma)}(y) = p_{\mathcal{N}\left(x,\frac{1}{2}t\right)}(y) = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{t}} exp\left\{-\frac{(y-x)^2}{t}\right\}$$

Fatta questa considerazione supponiamo di voler calcolare $\mathbb{E}\left[f(y) | y \sim \mathcal{N}(x, \frac{1}{2}t)\right]$

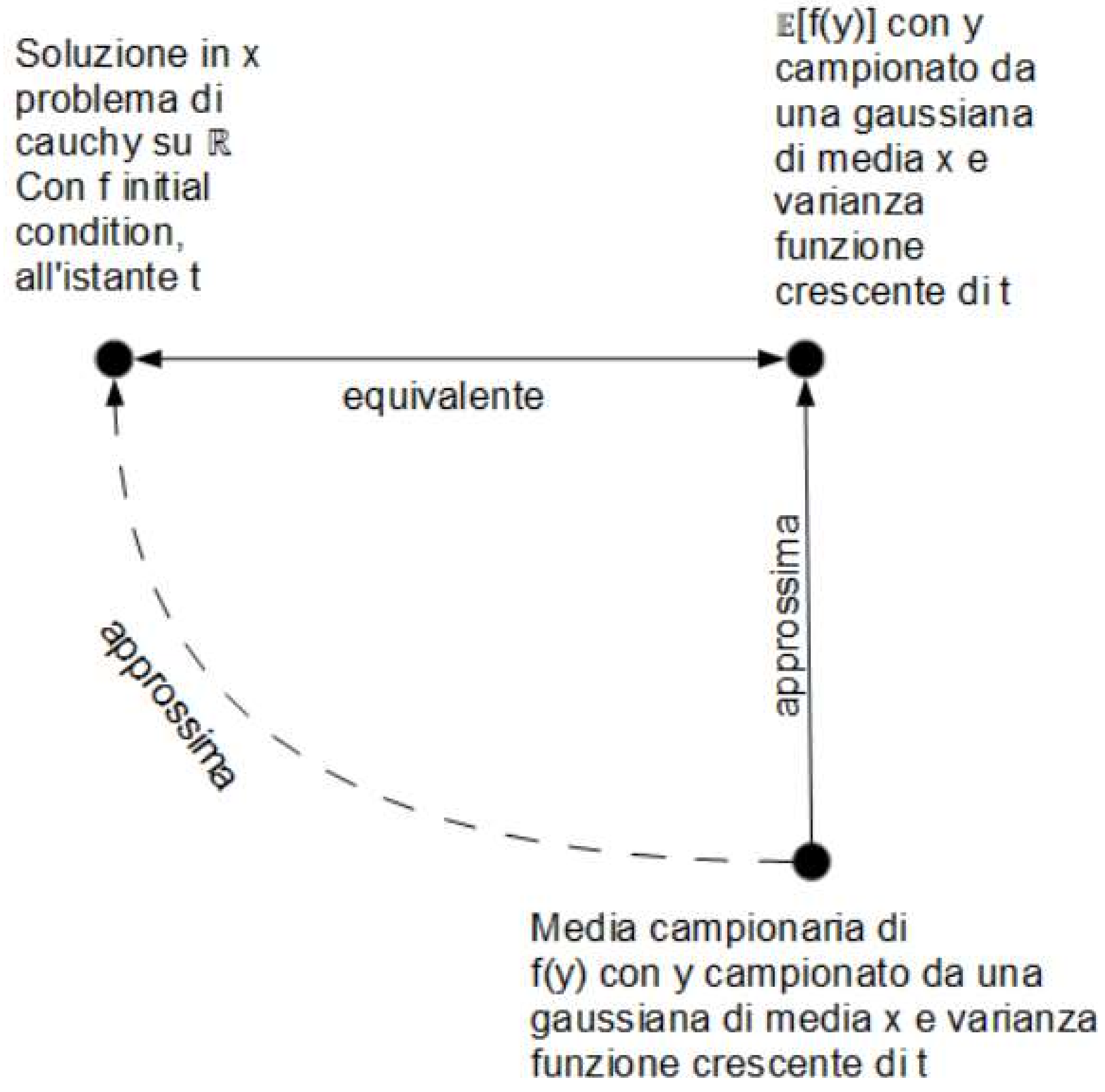
Dalla definizione di valore atteso abbiamo

$\mathbb{E}\left[f(y)|y \sim \mathcal{N}(x, \frac{1}{2}t)\right]$
$= \int_{\mathbb{R}} f(y) p_{\mathcal{N}\left(x, \frac{1}{2}t\right)}(y) dy$
$= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{t}} exp\left\{-\frac{(y-x)^2}{t}\right\}$
$= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{t}} \int_{\mathbb{R}} f(y) exp\left\{-\frac{(y-x)^2}{t}\right\}$
$= \frac{1}{\sqrt{\pi}} f_t(x)$

Sfruttiamo la legge dei grandi numeri:
Sia $Y_n := y_1, y_2, \ldots, y_n$ insieme di sample campionati dalla distribuzione $\mathcal{N}(x, \frac{1}{2}t)$
$lim_{n \to \infty} \quad \frac{1}{n} \sum_{y \in Y_n} f(y) = \mathbb{E}[f(y)|y \sim \mathcal{N}(x, \frac{1}{2}t)] = \frac{1}{\sqrt{\pi}} f_t(x)$

approssimazione

Sia $\bar{n}$ fissato il numero di campionamenti che quantifica la bontá della nostra approssimazione
Dunque :
$\frac{1}{\bar{n}} \sum_{y \in Y_{\bar{n}}} f(y) \approx \mathbb{E}[f(y)|y \sim \mathcal{N}(x, \frac{1}{2}t)] = \frac{1}{\sqrt{\pi}} f_t(x)$

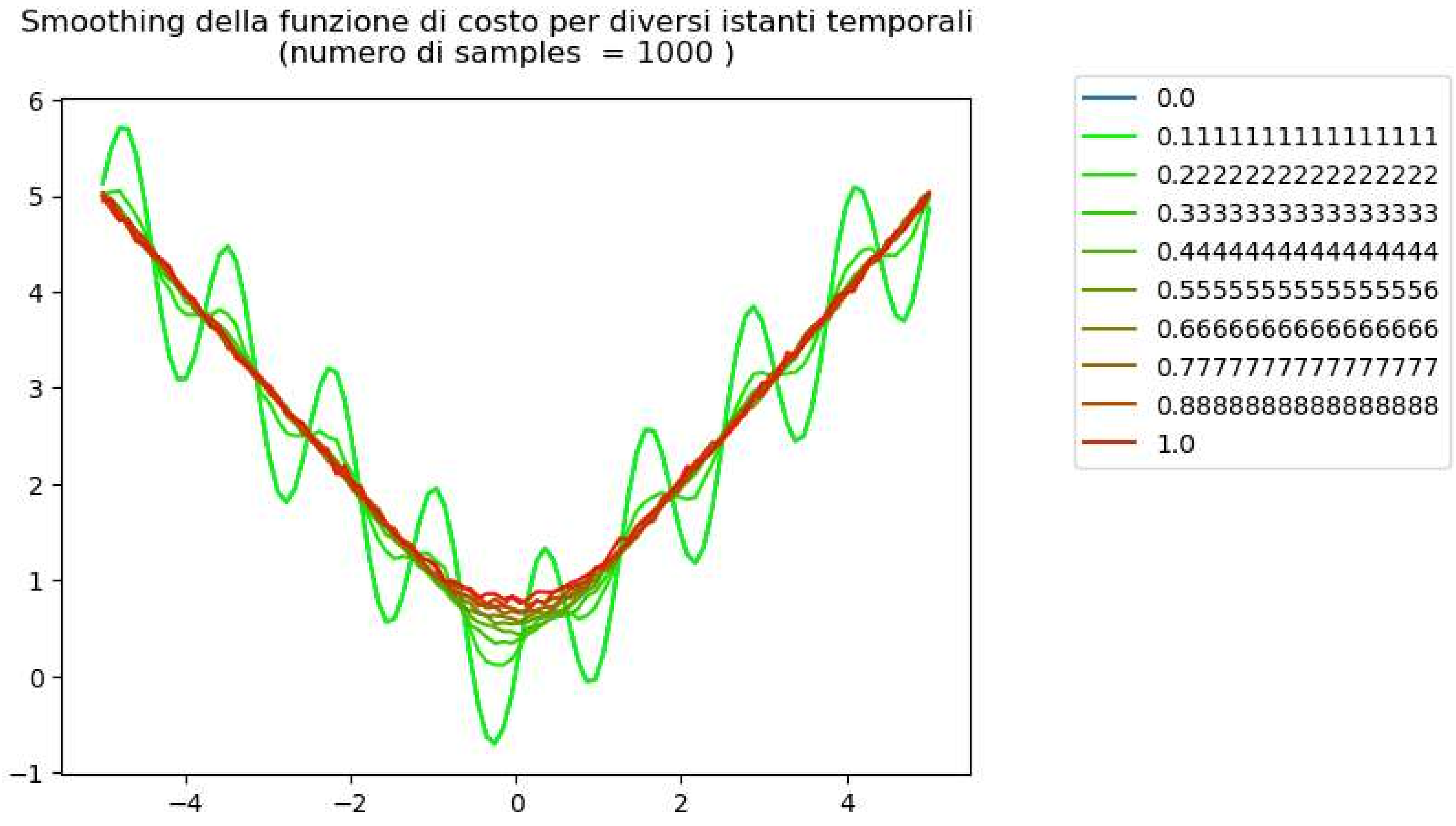


esperimento

In [217…

```
import numpy as np
import matplotlib.pyplot as plt

n_samples = 1000

def f(x):
    return np.sin(5 * x) + np.abs(x)

x_axis = np.linspace(-5,5,100)
plt.plot(x_axis,f(x_axis))

def relax(f,t):
    def r(x):
        Y = x + np.random.randn(n_samples) * np.sqrt(t)
        return f(Y).mean()
    return r

def color_grid(a, b, n):
    return np.array([
                np.array(a) * i / n + np.array(b) * ( n - i) / n
        for i in range(n)
    ])
```

```
n     = 10
cgrid = color_grid(
            [1.,0.,0.],
            [0.,1.,0.],
            n
)
tempi = np.linspace(0., 1., n)
for i,t in enumerate(tempi):
    res = np.array([ relax(f,t)(x) for x in x_axis ])
    plt.plot(x_axis, res, color = cgrid[i])
plt.legend( tempi , bbox_to_anchor=(1.1, 1.05))
plt.suptitle("Smoothing della funzione di costo per diversi istanti temporali \n (numero di samples  = %d )" % n_samples)
plt.show()
```



LA PROCEDURA DI OTTIMIZZAZIONE

inferenza della derivata della convoluzione stocastica

Vogliamo calcolare

$$\frac{\partial f_t(x)}{\partial x}$$

Supponendo che f abbia certe proprietá di regolaritá possiamo scrivere

$$\frac{\partial f_t(x)}{\partial x}$$

$$= \frac{\partial}{\partial x}\frac{1}{\sqrt{t}}\int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right)$$

$$= \frac{1}{\sqrt{t}}\int_{\mathbb{R}} f(y)\frac{\partial}{\partial x} exp\left(-\frac{(y-x)^2}{t}\right)$$

$$= \frac{1}{\sqrt{t}}\int_{\mathbb{R}} -2f(y) exp\left(-\frac{(y-x)^2}{t}\right)(y-x)$$

$$= \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right)(x-y)$$

$$= \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) xdy - \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) ydy$$

$$= \frac{2x}{\sqrt{t}}\int_{\mathbb{R}} f(y) exp\left(-\frac{(y-x)^2}{t}\right) dy - \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y)y\, exp\left(-\frac{(y-x)^2}{t}\right) dy$$

$$= 2xf_t(x) - 2g_t(x)\ with\ g(y) = f(y)y$$

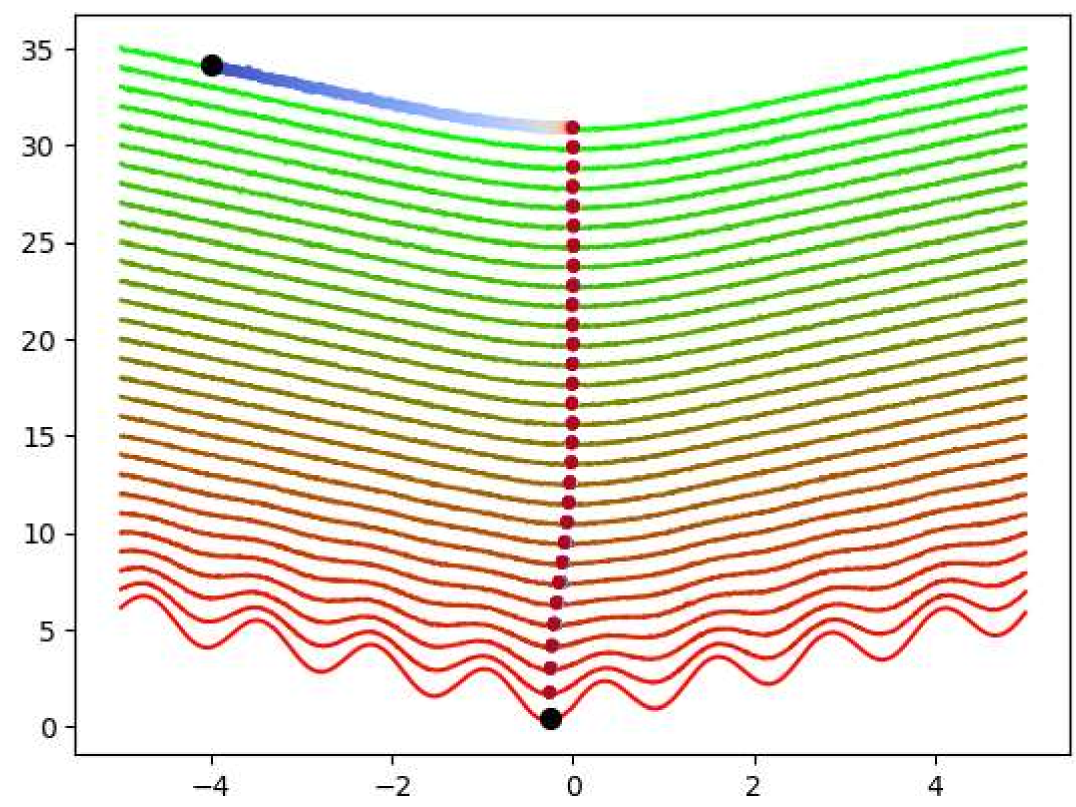
Applicazione della discesa del gradiente "a livelli"

In [225…

```
x    = -4.
x0   = x * 1.

def g(y):
    return f(y)*y

def relax_grad(f,t):
    def r(x):
        Y = x + np.random.randn(n_samples) * np.sqrt(t)
        return (2 *x * f(Y) - 2 * g(Y)) .mean()
    return r



n = 30
```

```
cgrid = color_grid(
            [1.,0.,0.],
            [0.,1.,0.],
            n
)

y0 = 0
x_axis = np.linspace(-5,5,1000)
for i,t in enumerate(np.linspace(0,1.,n)[-1::-1]):
    xs = list()
    ys = list()
    grad_r = relax_grad(f,t)
    f_r    = relax(f,t)
    mom_x = 0.
    for j in range(300):
        if i == 0:
            if j == 0:
                y0 = f_r(x)
        xs.append(x)
        ys.append(f_r(x))

        if t != 0:
            mom_x = 0.4 * mom_x + 0.01 * grad_r(x)
            x = x + mom_x
        else:
            h = 1e-5
            x = x + 1. * ( f(x + h) - f(x - h)  ) / 2 * h
    mom_x = 0.
    xs = np.array(xs)
    ys = np.array(ys)
    res = np.array([ f_r(x) for x in x_axis ])
    plt.plot(x_axis, res + n - i , color = cgrid[i])
    plt.plot   (xs,ys + 0.1 + n - i, color = 'black', lw = 3.)
    plt.scatter(xs,ys + 0.1 + n - i, c = np.arange(len(xs)), s = 10., zorder = 20000, cmap = 'coolwarm')


plt.scatter(xs[-1],ys[-1] + 0.1 + n - i, s = 50.,  zorder= 40000, color = 'black')
plt.scatter(x0 , y0  + n + 0.1, s = 50., zorder= 40000 , color = 'black')
plt.suptitle("Convergenza al minimo globale con x0 = %f" % x0)
plt.show()
```

Convergenza al minimo globale con x0 = -4.000000



Generalizzazione a piu variabili

Sia $f : \mathbb{R}^p \to \mathbb{R}$ una funzione di cui cerchiamo un punto di ottimo globale $\mathbf{x}^*$

Definiamo la funzione rilassata come

$f_t(\mathbf{x}) := \int_{\mathbb{R}^p} f(\mathbf{y})\Phi_k(\mathbf{y}-\mathbf{x},t)d\mathbf{y}$

Con $\Phi_k(\mathbf{z},t) := \frac{1}{\sqrt{(4\pi kt)^p}} exp\left(-\frac{\mathbf{z}^T\mathbf{z}}{4kt}\right)$

Poniamo per semplicitá $\hat{k} = \frac{1}{4\pi}$ **e otteniamo**

$\Phi(\mathbf{z},t) := \Phi_{\hat{k}}(\mathbf{z},t) = \frac{1}{\sqrt{t^p}} exp\left(-\frac{||\mathbf{z}||^2}{t}\right)$

Sostituendo in $f_t(x)$ **otteniamo**

$f_t(\mathbf{x}) = \int_{\mathbb{R}} f(\mathbf{y})\Phi(\mathbf{y}-\mathbf{x},t)dy$
$= \int_{\mathbb{R}} f(\mathbf{y})\frac{1}{\sqrt{t^p}}exp\left(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t}\right)$
$= \frac{1}{\sqrt{t^p}}\int_{\mathbb{R}} f(\mathbf{y})exp\left(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t}\right)$

Convoluzione stocastica in piu dimensioni

$$(2\pi)^{-k/2}\det(\boldsymbol{\Sigma})^{-1/2}\ \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

La densitá della distribuzione gaussiana multivariata $\mathcal{N}(\mu,\sigma^2 I)$ puo essere modellata come
$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}det(\sigma^2 I)^{-\frac{1}{2}}exp(-\frac{1}{2}(\mathbf{y}-\mu)^T[\sigma^2 I]^{-1}(\mathbf{y}-\mu))$
$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}\left(\sigma^{2p}det(I)\right)^{-\frac{1}{2}}exp(-\frac{1}{2}(\mathbf{y}-\mu)^T\sigma^{-2}I(\mathbf{y}-\mu))$
$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}\left(\sigma^{2p}\right)^{-\frac{1}{2}}exp(-\frac{1}{2}(\mathbf{y}-\mu)^T\sigma^{-2}(\mathbf{y}-\mu))$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}\left(\sigma^{2p}\right)^{-\frac{1}{2}}exp(-\frac{1}{2}\frac{(\mathbf{y}-\mu)^T(\mathbf{y}-\mu)}{\sigma^2})$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}\left(\sigma^{2p}\right)^{-\frac{1}{2}}exp(-\frac{1}{2}\frac{||\mathbf{y}-\mu||^2}{\sigma^2})$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}}\left(\sigma^{2}\right)^{-\frac{p}{2}}exp(-\frac{1}{2}\frac{||\mathbf{y}-\mu||^2}{\sigma^2})$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = \left(2\pi\sigma^{2}\right)^{-\frac{p}{2}}exp(-\frac{1}{2}\frac{||\mathbf{y}-\mu||^2}{\sigma^2})$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = \left(2\pi\sigma^{2}\right)^{-\frac{p}{2}}exp(-\frac{||\mathbf{y}-\mu||^2}{2\sigma^2})$

$p_{\mathcal{N}(\mu,\sigma^2I)}(\mathbf{y}) = (\pi)^{-\frac{p}{2}}\left(2\sigma^{2}\right)^{-\frac{p}{2}}exp(-\frac{||\mathbf{y}-\mu||^2}{2\sigma^2})$

Se nomino
$t := 2\sigma^2 \implies \sigma^2 = \frac{1}{2}t$
$\mathbf{x} := \mu \implies \mu = \mathbf{x}$

$p_{\mathcal{N}(\mathbf{x},\frac{1}{2}tI)}(\mathbf{y}) = (\pi)^{-\frac{p}{2}}(t)^{-\frac{p}{2}}exp(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t})$

$p_{\mathcal{N}(\mathbf{x},\frac{1}{2}tI)}(\mathbf{y}) = (\pi)^{-\frac{p}{2}}\frac{1}{(\sqrt{t^p})}exp(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t})$

Fatta questa considerazione supponiamo di voler calcolare $\mathbb{E}\left[f(y)|y\sim\mathcal{N}(x,\frac{1}{2}t)\right]$
Dalla definizione di valore atteso abbiamo
$\mathbb{E}\left[f(\mathbf{y})|\mathbf{y}\sim\mathcal{N}(\mathbf{x},\frac{1}{2}tI)\right]$
$= \int_{\mathbb{R}} f(\mathbf{y})p_{\mathcal{N}\left(\mathbf{x},\frac{1}{2}tI\right)}(\mathbf{y})d\mathbf{y}$
$= \int_{\mathbb{R}} f(\mathbf{y})(\pi)^{-\frac{p}{2}}\frac{1}{(\sqrt{t^p})}exp(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t})$
$= \int_{\mathbb{R}} f(\mathbf{y})(\pi)^{-\frac{p}{2}}\frac{1}{(\sqrt{t^p})}exp(-\frac{||\mathbf{y}-\mathbf{x}||^2}{t})$
$= (\pi)^{-\frac{p}{2}}f_t(\mathbf{x})$

Vogliamo calcolare
$\frac{\partial f_t(\mathbf{x})}{\partial x_i}$
Supponendo che f abbia certe proprietá di regolaritá possiamo scrivere
$\frac{\partial f_t(\mathbf{x})}{\partial x_i}$
$= \frac{\partial}{\partial x_i}\frac{1}{\sqrt{t}}\int_{\mathbb{R}} f(y)exp\left(-\frac{(y-x)^2}{t}\right)$
$= \frac{1}{\sqrt{t}}\int_{\mathbb{R}} f(y)\frac{\partial}{\partial x}exp\left(-\frac{(y-x)^2}{t}\right)$
$= \frac{1}{\sqrt{t}}\int_{\mathbb{R}} -2f(y)exp\left(-\frac{(y-x)^2}{t}\right)(y-x)$
$= \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y)exp\left(-\frac{(y-x)^2}{t}\right)(x-y)$
$= \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y)exp\left(-\frac{(y-x)^2}{t}\right)xdy - \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y)exp\left(-\frac{(y-x)^2}{t}\right)ydy$
$= \frac{2x}{\sqrt{t}}\int_{\mathbb{R}} f(y)exp\left(-\frac{(y-x)^2}{t}\right)dy - \frac{2}{\sqrt{t}}\int_{\mathbb{R}} f(y)y\,exp\left(-\frac{(y-x)^2}{t}\right)dy$
$= 2xf_t(x) - 2g_t(x)\ with\ g(y) = f(y)y$