From Statistical Estimation and Generative Modeling to Inverse Problems

LECTURE 5: INVERSE PROBLEMS

Inverse Problems

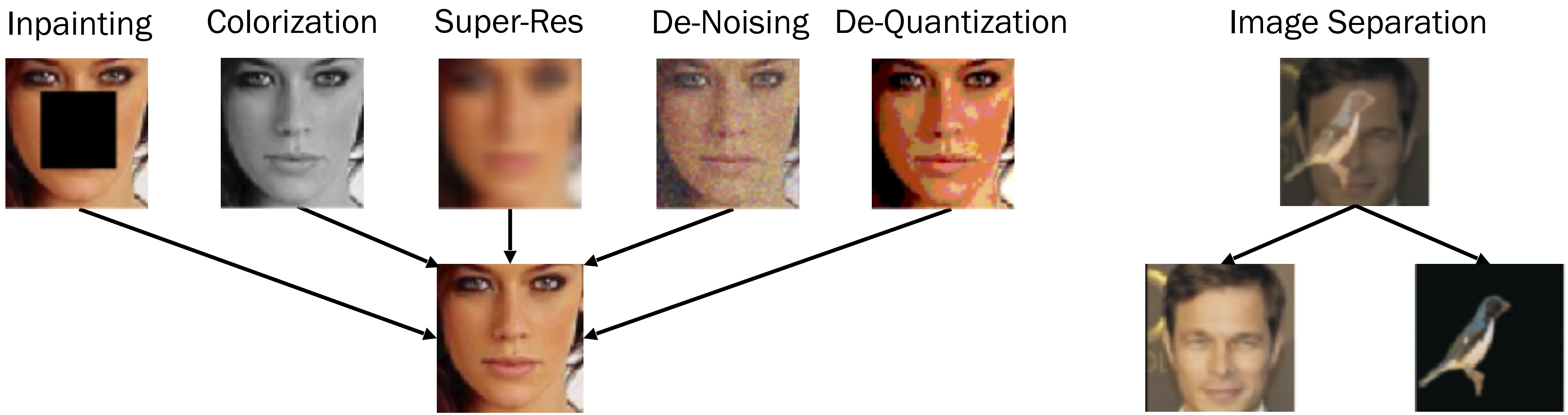
- Problem Definition
- Early Efforts
- Solutions Based on GANs
  - Add-Hoc
  - Applying Statistical Estimation
- Examples

Problem Definition

We are given vector X (measurements) and interested in estimating vector Y

We assume $X = \mathsf{T}(Y) + W$ where $\mathsf{T}(Y)$ general (mostly) known transformation

Basic characteristic: $\boxed{\dim(X) \;\leq\; \dim(Y)}$



Early Efforts

<u>Inpainting</u>

Diffusion based inpainting

No prior information, training data not very useful

Solutions Based on GANs

Available training data $\{X_1,\dots,X_n\}$

Design generator $\mathsf{G}(Z)$ so that when applied to Z with $Z \sim \mathsf{h}(Z)$ then $Y=\mathsf{G}(Z)$ has same density as training data

Generate $\{Z_1,\dots,Z_m\}$ with density $\mathsf{h}(Z)$

$$\hat{\mathsf{J}}(\theta,\vartheta)=\frac{1}{n}\sum_{t=1}^{n}\phi\big(\underbrace{\mathsf{D}(X_t,\vartheta)}_{\text{Discriminator}}\big)+\frac{1}{m}\sum_{t=1}^{m}\psi\Big(\mathsf{D}\big(\underbrace{\mathsf{G}(Z_t,\theta)}_{\text{Generator}},\vartheta\big)\Big)$$

$$\min_{\theta}\max_{\vartheta}\hat{\mathsf{J}}(\theta,\vartheta)$$

Assume known Generative Model $\{\mathsf{G}(Z),\mathsf{h}(Z)\}$, also Discriminator $\mathsf{D}(X)$

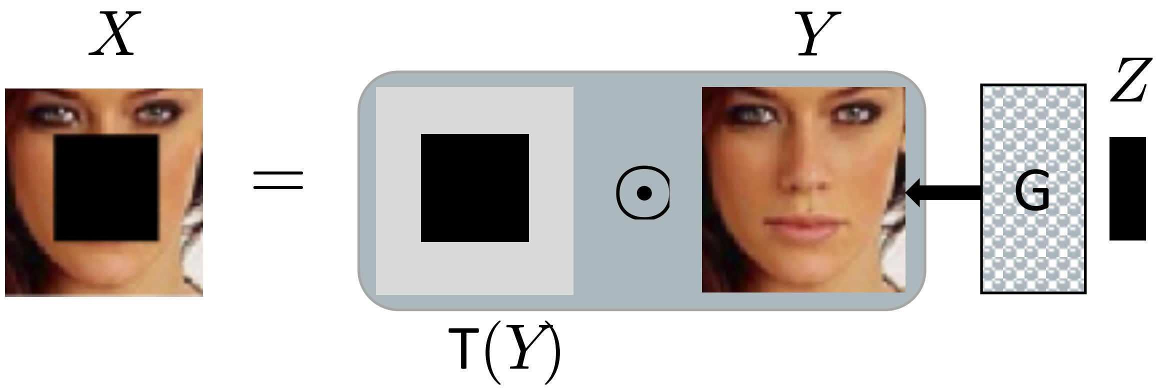
General Problem

Given vector X (measurements) we are interested in estimating vector Y with $X = \mathsf{T}(Y) + W$ and $\mathsf{T}(Y)$ known transformation

Y follows Generative Model

There exists Z following density $\mathsf{h}(Z)$ such that $\quad Y = \mathsf{G}(Z)$

Instead of estimating Y from X we estimate input to generative model Z



Measurements $X = \mathsf{T}(Y) + W$

Compute estimate $\hat{Y}$

Measurements $X = \mathsf{T}\big(\mathsf{G}(Z)\big) + W$

Compute estimate $\hat{Z}$

and let $\hat{Y} = \mathsf{G}(\hat{Z})$

Since $\dim(Z) \ll \dim(Y)$

Significant computational gain and more stable processing

Select $\hat{Z}$ so that measurements X and $\mathsf{T}\big(\mathsf{G}(\hat{Z})\big)$ are "close"

$$\min_{Z} \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\|^2 \;\Rightarrow\; \hat{Z} = \arg\min_{Z} \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\|^2$$

Well defined optimization, computationally stable

Doesn't Work!!

$\hat{Y} = \mathsf{G}(\hat{Z})$ does not have the correct characteristics

Recall that generative model is the pair $\{\mathsf{G}(Z), \mathsf{h}(Z)\}$

Even if $\mathsf{T}\big(\mathsf{G}(\hat{Z})\big)$ "close" to X, if likelihood $\mathsf{h}(\hat{Z}) \ll 1$

then $\hat{Y} = \mathsf{G}(\hat{Z})$ is "bad" result

Must take into account input density $\mathsf{h}(Z)$

Ad-Hoc Solutions

Yeh et al. (2017), (2018)

Assumes generative model trained with cross entropy method

$\phi(z) = \log(1-z),\ \psi(z) = \log(z),\ z \in (0,1)\ \Rightarrow$

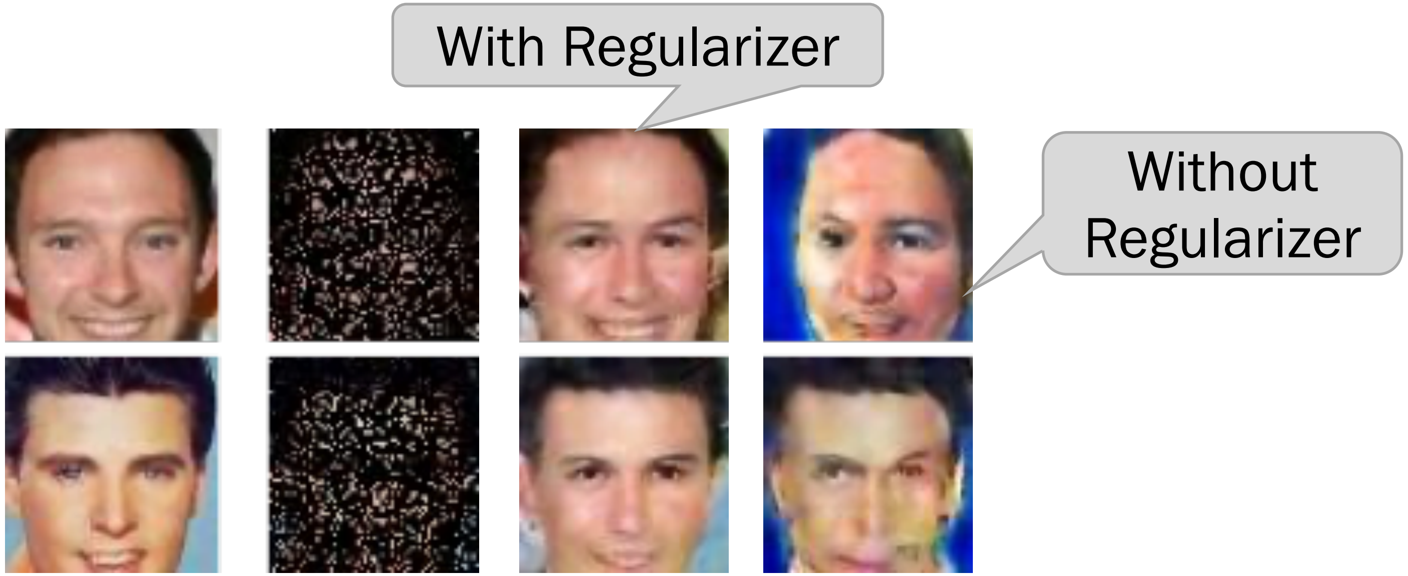
Design $\mathsf{G}(Z)$ generator, $\mathsf{D}(Z)$ discriminator, when $Z \sim \mathsf{h}(Z)$

Regularizer

$$\mathsf{J}(Z) = \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\|_{L_1} + \lambda \left\{ \log\Big(1 - \mathsf{D}\big(\mathsf{G}(Z)\big)\Big) - \log\Big(\mathsf{D}\big(\mathsf{G}(Z)\big)\Big) + \log\big(\mathsf{h}(Z)\big) \right\}$$

Parameter
needs tuning

$$\hat{Z} = \arg\min_{Z} \mathsf{J}(Z), \quad \hat{Y} = \mathsf{G}(\hat{Z})$$



To tune parameter λ need ideal images $\{Y_1,\ldots,Y_n\}$ apply transformation $X_i=\mathsf{T}(Y_i)$

For various values of λ for each $\{X_1,\ldots,X_n\}$ compute

$$\hat{Z}_1(\lambda),\ldots,\hat{Z}_n(\lambda) \;\Rightarrow\; \hat{Y}_1(\lambda),\ldots,\hat{Y}_n(\lambda)$$

$$\lambda_i = \arg\min_{\lambda} \|Y_i - \hat{Y}_i(\lambda)\|$$

$$\lambda = \frac{1}{n}\{\lambda_1 + \cdots + \lambda_n\}$$

Requires exact knowledge of $\mathsf{T}(Y)$

Asim et al. (2019)

$$\hat{Z} = \arg\min_{Z} \left\{ \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\|^2 - \lambda \log\big(\mathsf{h}(Z)\big) \right\}$$

Parameter needs tuning

Requires exact knowledge of $\mathsf{T}(Y)$

Siavelis et al. (2020)

$$\hat{Z} = \underset{Z}{\text{arg min}} \left\{ \|X - \mathsf{T}(\mathsf{G}(Z))\|_{L_1} - \lambda\Big(1 - \mathsf{D}(\mathsf{G}(Z))\Big) \right\}$$

Parameter needs tuning

Requires exact knowledge of $\mathsf{T}(Y)$

Applying Statistical Estimation

Given densities $\mathsf{f}(X|Z)$ and prior $\mathsf{h}(Z)$, for measurement X estimate Z

MAP estimator is of the form

$$\hat{Z} = \arg\max_{Z} \mathsf{f}(Z|X) = \arg\max_{Z} \mathsf{f}(X, Z) = \arg\max_{Z} \mathsf{f}(X|Z)\mathsf{h}(Z)$$

Let $Z=\{Z_1,Z_2\}$ where there is prior for Z_1 but not for Z_2

Treat non-existing prior as degenerate uniform

$$\begin{aligned}
\arg\max_{Z_1,Z_2} \mathsf{f}(Z_1, Z_2|X) &= \arg\max_{Z_1,Z_2} \mathsf{f}(X, Z_1, Z_2) \\
&= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}(Z_1, Z_2) \\
&= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1|Z_2)\mathsf{h}_2(Z_2) \\
\mathsf{h}_2(Z_2) \text{ degenerate uniform} \quad &= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1|Z_2)
\end{aligned}$$

If interested in estimating Z_1 and Z_2 are nuisance parameters then

$$\hat{Z}_1 = \arg\max_{Z_1} \left\{ \max_{Z_2} \mathrm{f}(X|Z_1, Z_2)\mathrm{h}_1(Z_1|Z_2) \right\}$$

where $\mathrm{h}_1(Z_1|Z_2)$ prior of Z_1 given Z_2

If Z_1 does not depend on Z_2 then $\mathrm{h}_1(Z_1|Z_2)=\mathrm{h}_1(Z_1)$ and

$$\boxed{\hat{Z}_1 = \arg\max_{Z_1} \left\{ \max_{Z_2} \mathrm{f}(X|Z_1, Z_2)\mathrm{h}_1(Z_1) \right\}}$$

We are given vector X (measurements) and interested in estimating vector Y

We assume $X = \mathsf{T}(Y, \alpha) + W = \mathsf{T}(\mathsf{G}(Z), \alpha) + W$

$\mathsf{T}(Y, \alpha)$: transformation of known mathematical form possibly containing unknown parameters α. Can be different per measurement X

W: additive noise with density $\mathsf{g}_w(W, \beta)$ possibly containing unknown parameters β. Can be different per measurement X

Z: follows density $\mathsf{h}(Z)$ from generative model $\{\mathsf{G}(Z), \mathsf{h}(Z)\}$

$Z_1 = Z$, $Z_2 = \{\alpha, \beta\}$ $\quad \mathsf{f}(X|Z_1, Z_2) = \mathsf{f}(X|Z, \alpha, \beta) = \mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z), \alpha\big), \beta\Big)$

$$\hat{Z} = \arg\max_{Z} \left\{ \max_{\alpha,\beta} \mathsf{f}(X|Z, \alpha, \beta)\mathsf{h}(Z) \right\}$$

$$= \arg\max_{Z} \left\{ \max_{\alpha,\beta} \mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z), \alpha\big), \beta\Big)\mathsf{h}(Z) \right\}$$

$$\hat{Z} = \text{arg}\max_{Z}\left\{\max_{\alpha}\max_{\beta}\mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z),\alpha\big),\beta\Big)\mathsf{h}(Z)\right\}$$

W: additive noise is Gaussian mean 0 and covariance $\beta^2 I$

$$\max_{\beta}\mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z),\alpha\big),\beta\Big) = \max_{\beta}\frac{e^{-\|X-\mathsf{T}(\mathsf{G}(Z),\alpha)\|^2/2\beta^2}}{(\sqrt{2\pi\beta^2})^N}$$

$$= \frac{\mathsf{C}}{\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^N}$$

N: length of measurement vector X

$$\hat{Z} = \text{arg}\max_{Z}\left\{\max_{\alpha}\frac{\mathsf{C}\,\mathsf{h}(Z)}{\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^N}\right\}$$

$$= \text{arg}\max_{Z}\frac{\mathsf{h}(Z)}{(\min_{\alpha}\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^2)^{N/2}}$$

Z: If input of generative model is Gaussian with mean 0 and covariance identity

$$\hat{Z} = \arg\max_{Z} \frac{\mathsf{C}'\, e^{-\|Z\|^2/2}}{(\min_\alpha \|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2)^{N/2}}$$

$$\boxed{\hat{Z} = \arg\min_{Z} \left\{ \log\left(\min_{\alpha} \|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2 \right) + \frac{1}{N}\|Z\|^2 \right\}}$$

$$\Leftrightarrow \quad \min_{Z,\alpha} \left\{ \log\left(\|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2 \right) + \frac{1}{N}\|Z\|^2 \right\}$$

If transformation satisfies

$$\mathsf{T}(Y,\alpha) = \alpha_1 \mathsf{T}_1(Y) + \cdots + \alpha_m \mathsf{T}_m(Y)$$

then $\mathsf{T}\big(\mathsf{G}(Z),\alpha\big) = \alpha_1 \mathsf{T}_1\big(\mathsf{G}(Z)\big) + \cdots + \alpha_m \mathsf{T}_m\big(\mathsf{G}(Z)\big) = \mathcal{G}(Z)A$

where $\mathcal{G}(Z) = \Big[\mathsf{T}_1\big(\mathsf{G}(Z)\big) \cdots \mathsf{T}_m\big(\mathsf{G}(Z)\big)\Big], \;\; A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}$

$$\min_A \|X - \mathcal{G}(Z)A\|^2 = \|X\|^2 - X^\intercal \mathcal{G}(Z)\big(\mathcal{G}^\intercal(Z)\mathcal{G}(Z)\big)^{-1}\mathcal{G}^\intercal(Z)X$$

$$\hat{Z} = \arg\min_Z \left\{ \log\Big(\|X\|^2 - X^\intercal \mathcal{G}(Z)\big(\mathcal{G}^\intercal(Z)\mathcal{G}(Z)\big)^{-1}\mathcal{G}^\intercal(Z)X\Big) + \frac{1}{N}\|Z\|^2 \right\}$$

Examples

Blurring with 3 X 3 kernel

Mth 1 Mth 2 Known Unknown

Colorization (green channel)

Mth 1 Mth 2 Known Unknown

No parameters to tune

De-Quantization

2 levels per RGB channel, 8 colors

De-Quantization

3 levels per RGB channel, 27 colors

De-Quantization

5 levels per RGB channel, 125 colors

De-Quantization and Colorization

RGB → Gray → BW (2 levels)

With threshold estimation

Data Mixtures

We assume two independent data vectors Y_1, Y_2 that are combined as

$$X = \alpha_1 Y_1 + \alpha_2 Y_2 + W$$

From mixture X recover original Y_1, Y_2

Y_1: generative model {$\mathsf{G}_1(Z_1)$,$\mathsf{h}_1(Z_1)$}

Y_2: generative model {$\mathsf{G}_2(Z_2)$,$\mathsf{h}_2(Z_2)$}

W: additive noise with density $\mathsf{g}_w(W,\beta)$ containing unknown parameters β

If W Gaussian, mean 0, covariance $\beta^2 I$

If inputs of both generative models Gaussian, mean 0, covariance identity

known

$$\{\hat{Z}_1, \hat{Z}_2\} = \arg\min_{Z_1, Z_2} \left\{ \log\left(\|X - \alpha_1 \mathsf{G}_1(Z_1) - \alpha_2 \mathsf{G}_2(Z_2)\|^2\right) + \frac{1}{N}\left(\|Z_1\|^2 + \|Z_2\|^2\right) \right\}$$

unknown

$$\{\hat{Z}_1, \hat{Z}_2\} = \arg\min_{Z_1, Z_2} \left\{ \log\Big(\min_{\alpha_1, \alpha_2} \|X - \alpha_1 \mathsf{G}_1(Z_1) - \alpha_2 \mathsf{G}_2(Z_2)\|^2 \Big) + \frac{1}{N}\big(\|Z_1\|^2 + \|Z_2\|^2\big) \right\}$$

$$= \arg\min_{Z_1, Z_2} \Big\{ \log\Big(\|X\|^2 - X^\mathsf{T} \mathcal{G}(Z_1, Z_2) \Big(\mathcal{G}^\mathsf{T}(Z_1, Z_2) \mathcal{G}(Z_1, Z_2) \Big)^{-1} \mathcal{G}^\mathsf{T}(Z_1, Z_2) X \Big) + \frac{1}{N}\big(\|Z_1\|^2 + \|Z_2\|^2\big) \Big\}$$

where $\mathcal{G}(Z_1, Z_2) = \big[\mathsf{G}_1(Z_1)\ \mathsf{G}_2(Z_2) \big]$


Known coefs
Uknown coefs


Known coefs
Uknown coefs

Nonlinear Data Mixtures

We assume two independent data vectors Y_1, Y_2 that are combined as

$$X = \mathsf{T}(Y_1, Y_2, \alpha) + W$$

Y_1: generative model $\{\mathsf{G}_1(Z_1), \mathsf{h}_1(Z_1)\}$

Y_2: generative model $\{\mathsf{G}_2(Z_2), \mathsf{h}_2(Z_2)\}$

W: additive noise with density $\mathsf{g}_w(W, \beta)$ containing unknown parameters β

$$\{\hat{Z}_1, \hat{Z}_2\} =$$

$$\arg \min_{Z_1, Z_2} \left\{ \log \left(\min_{\alpha} \|X - \mathsf{T}\big(\mathsf{G}_1(Z_1), \mathsf{G}_2(Z_2), \alpha\big)\|^2 \right) + \frac{1}{N}\left(\|Z_1\|^2 + \|Z_2\|^2\right) \right\}$$

$$\Rightarrow \min_{Z_1, Z_2, \alpha} \left\{ \log \left(\|X - \mathsf{T}\big(\mathsf{G}_1(Z_1), \mathsf{G}_2(Z_2), \alpha\big)\|^2 \right) + \frac{1}{N}\left(\|Z_1\|^2 + \|Z_2\|^2\right) \right\}$$