Lecture 28: Generative Plug-and-Play

Charles A. Bouman

Purdue University

Friday, October 30, 2026

Outline

1. Why sample instead of optimize
2. A denoiser is a score estimate
3. From proximal map to proximal generator: the GPnP algorithm
4. The connection to diffusion models, and a warning

Reading: C. A. Bouman and G. T. Buzzard, "Generative Plug and Play: Posterior Sampling for Inverse Problems," Allerton Conference, 2023.

Today's Big Ideas

- Samples show which features the data supports; one image hides this.
- A Gaussian-noise MMSE denoiser gives the score of the blurred prior.
- A proximal generator draws a sample from the proximal distribution, and alternating the prior and forward proximal generators is a Markov chain whose stationary distribution is the posterior with a blurred prior.
- In practice, the prior generator is a denoiser step plus noise, the forward generator is the proximal map plus noise, and the noise level is annealed from large to small.
- With no data, GPnP is annealed Langevin dynamics, and wherever the data does not constrain the image, the samples come from the prior.

PnP returns one reconstruction

- PnP-ADMM and consensus equilibrium compute one image x^* from the data y, a forward model, and a denoiser.
- With few or noisy measurements, many images fit the data.
- One image does not show which features the data supports and which features the prior invented.

Goal

Generate random samples $X^{(1)}, \ldots, X^{(K)}$ from the posterior distribution $p_{x|y}(x|y)$. The pixelwise sample mean and sample standard deviation of these samples show the uncertainty in the reconstruction.

The posterior distribution in terms of energy functions

Define the prior and forward **energy functions** as

$$u_0(x) = -\log p_0(x) + C_0 \ ,$$
$$u_1(x) = -\log p_{y|x}(y|x) + C_1 \ ,$$

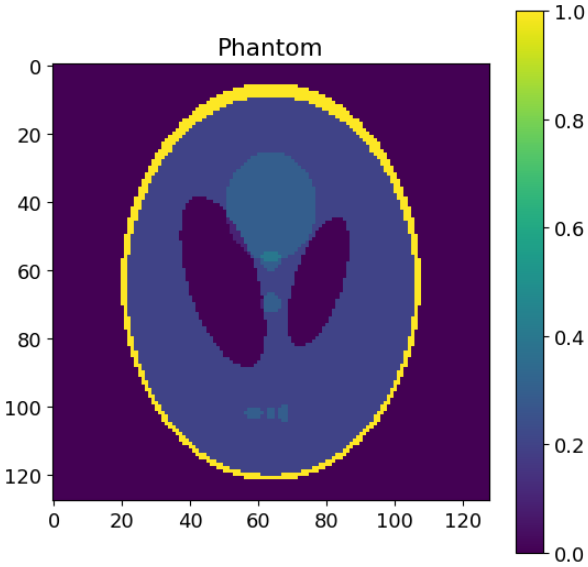
where $p_0(x)$ is the prior distribution and $p_{y|x}(y|x)$ is the forward model.

By Bayes' rule, the posterior distribution is

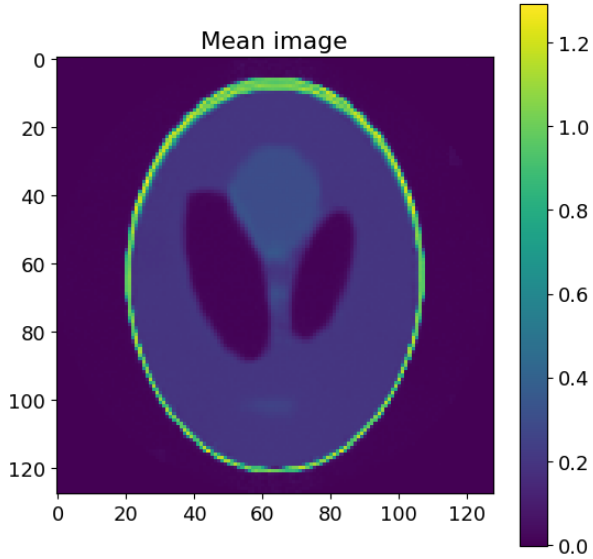
$$p_{x|y}(x|y) = \frac{1}{Z} \exp\left\{-u_1(x) - u_0(x)\right\} \ .$$

Generative Plug-and-Play (GPnP) generates samples from $p_{x|y}(x|y)$. Like PnP, it uses only two modules: a physics-based forward model and an image denoiser.

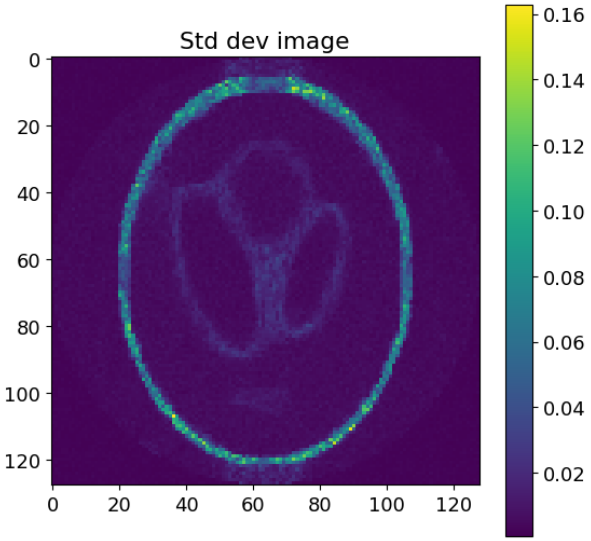
Sparse-view CT: posterior mean and standard deviation



Ground truth



Mean of 10 samples



Std. dev. of 10 samples

- The phantom is 128×128, with 16 views and a BM3D prior.
- The standard deviation is largest at the edges of the phantom.

The score of a distribution

The **score** of a density $p(x) = \frac{1}{Z} \exp\{-u(x)\}$ is defined as

$$s(x) = \nabla_x \log p(x) = -\nabla_x u(x) \ .$$

- The score does not depend on the normalizing constant Z.
- The score points where $\log p(x)$ increases fastest.
- The score determines $u(x)$ up to an additive constant.

The prior blurred by Gaussian noise

Let $X \sim p_0(x)$, let $W \sim N(0, I)$ be independent of X, and let $\tilde{X} = X + \sigma W$. Then $\tilde{X}$ has the **blurred prior** density

$$\tilde{p}_{0,\sigma^2}(x) = (p_0 * g_{\sigma^2})(x) \ , \qquad g_{\sigma^2}(x) = \frac{1}{(2\pi\sigma^2)^{p/2}} \exp\left\{ -\frac{1}{2\sigma^2} \|x\|^2 \right\} \ ,$$

where $*$ is convolution and p is the number of pixels.

Define the blurred energy function and its score as

$$\tilde{u}_0(x; \sigma^2) = -\log \left(e^{-u_0} * g_{\sigma^2} \right)(x) \ ,$$
$$s(x; \sigma^2) = -\nabla_x \tilde{u}_0(x; \sigma^2) \ .$$

- $s(x; \sigma^2)$ is the score of $\tilde{p}_{0,\sigma^2}$, and $\tilde{p}_{0,\sigma^2} \to p_0$ as $\sigma \to 0$.

Tweedie's formula

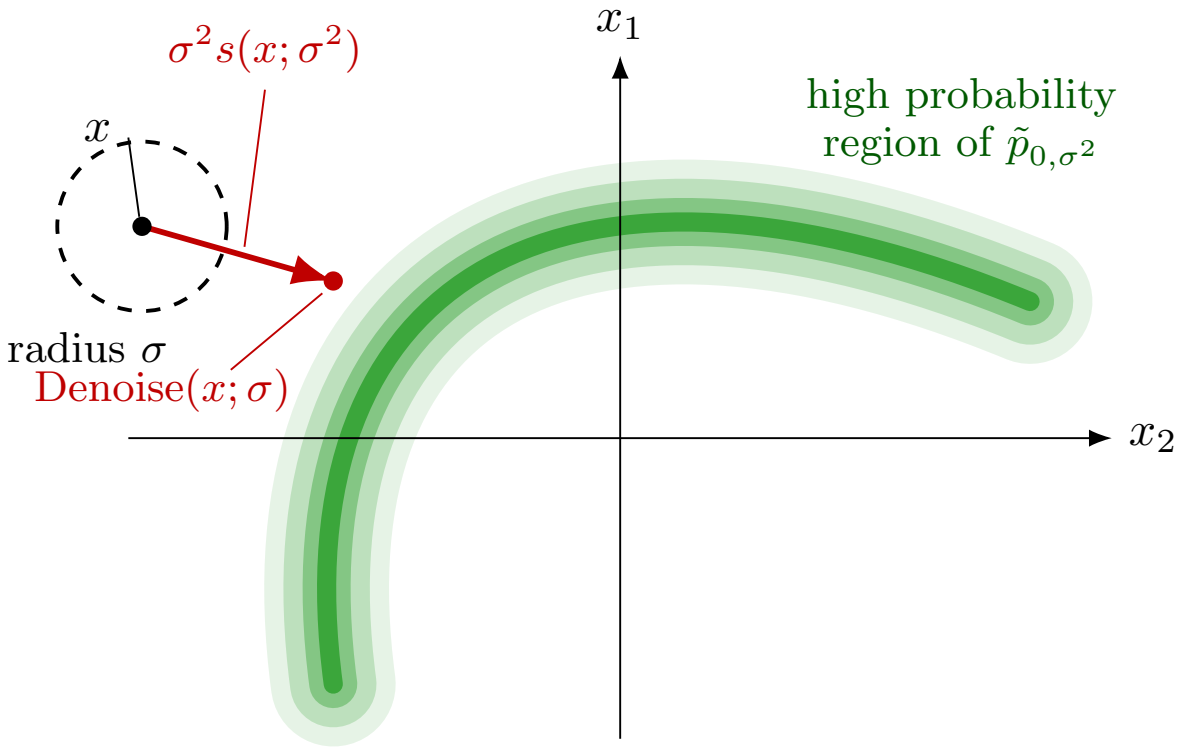
Property (Tweedie's formula; P. Vincent, 2011)

Let $\text{Denoise}(x;\sigma) = \mathbb{E}[X \mid \tilde{X} = x]$ be the MMSE denoiser for white Gaussian noise of variance σ^2. Then

$$\text{Denoise}(x;\sigma) = x + \sigma^2 s(x;\sigma^2) \ , \qquad s(x;\sigma^2) = \frac{1}{\sigma^2}\left[\text{Denoise}(x;\sigma) - x\right] \ .$$

- The formula is exact for every σ.
- It is the score of $\tilde{p}_{0,\sigma^2}$, not of p_0.
- An MMSE-trained denoiser estimates the score.

A denoiser moves a point toward higher probability



- The denoiser moves x by $\sigma^2 s(x;\sigma^2)$ uphill.
- Larger σ spreads the blurred prior; the score reaches farther.

Proximal distributions and proximal generators

For $i = 0, 1$, define the **proximal distributions**

$$q_i(x|v) = \frac{1}{Z_i(v)} \exp\left\{-u_i(x) - \frac{1}{2\gamma^2}\|x - v\|^2\right\} \ ,$$

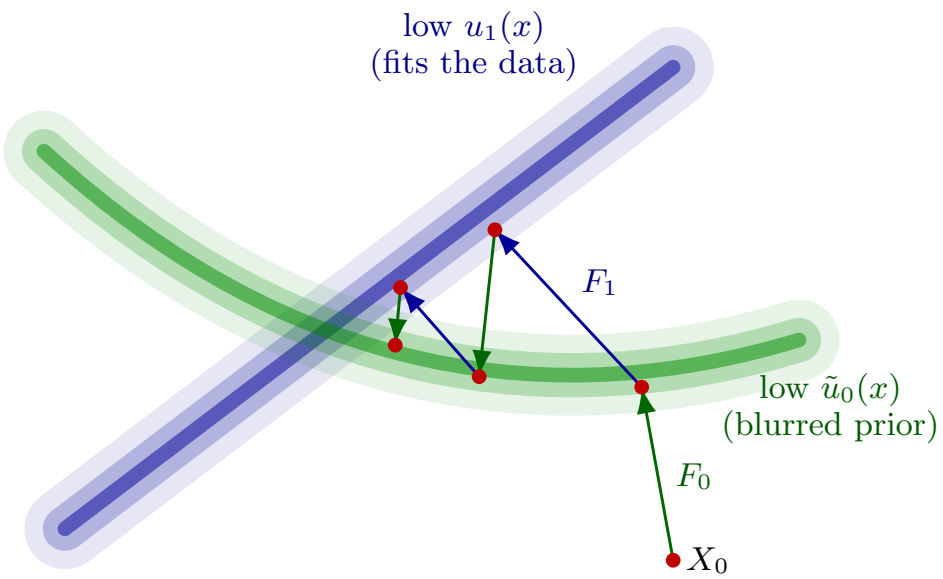
where $\gamma > 0$ is a parameter and $Z_i(v)$ is a normalizing constant.

The **proximal generator** $F_i(V)$ is a new random vector with conditional density $q_i(x|V)$, conditionally independent of all earlier random vectors given V.

Proximal map versus proximal generator

The proximal map $\bar{F}_i(v) = \arg\min_x \left\{u_i(x) + \frac{1}{2\gamma^2}\|x - v\|^2\right\}$ is the mode of $q_i(x|v)$. The proximal generator draws a sample from $q_i(x|v)$.

The GPnP Markov chain



GPnP alternates the prior generator F_0 and the forward generator F_1:

$$X_n = F_1(F_0(X_{n-1})) \ .$$

- F_0 moves toward high prior.
- F_1 moves toward a good fit.
- Each step adds randomness, so X_n never settles at one point.

X_n is a Markov chain. The question is which distribution it samples.

Theorem 1: the stationary distribution of GPnP

Theorem 1 (Bouman and Buzzard, 2023)

Let $X_n = F_1(F_0(X_{n-1}))$. Then X_n is a reversible Markov chain with stationary distribution

$$\tilde{p}_{\gamma^2}(x) = \frac{1}{Z'} \exp\left\{-u_1(x) - \tilde{u}_0(x;\gamma^2)\right\} \ ,$$

where $\tilde{u}_0(x;\gamma^2) = -\log\left(e^{-u_0} * g_{\gamma^2}\right)(x)$.

- $\tilde{p}_{\gamma^2}$ is the posterior with the prior blurred by a Gaussian of variance γ^2. As $\gamma \to 0$, $\tilde{p}_{\gamma^2}(x) \to p_{x|y}(x|y)$.
- If ergodic, X_n converges in distribution to $\tilde{p}_{\gamma^2}$.

A density π is a **stationary distribution** of the chain if $X_{n-1} \sim \pi$ implies $X_n \sim \pi$. The chain is **reversible** if, when $X_{n-1} \sim \pi$, the pair (X_{n-1}, X_n) has the same distribution as (X_n, X_{n-1}).

Why Theorem 1 holds

Define the joint density of a pair (x_0, x_1) as

$$p(x_0, x_1) = \frac{1}{Z} \exp\left\{-u_1(x_1) - u_0(x_0) - \frac{1}{2\gamma^2}\|x_0 - x_1\|^2\right\} \ .$$

- Its conditionals are $p(x_0|x_1) = q_0(x_0|x_1)$ and $p(x_1|x_0) = q_1(x_1|x_0)$.
- So the update $x_0 \leftarrow F_0(x_1)$, $x_1 \leftarrow F_1(x_0)$ is a two-block Gibbs sampler with stationary distribution $p(x_0, x_1)$.
- Integrating out x_0 turns the coupling into a Gaussian blur:

$$\begin{aligned}\int e^{-u_0(x_0) - \frac{1}{2\gamma^2}\|x_0 - x_1\|^2} dx_0 &= (2\pi\gamma^2)^{p/2} \left(e^{-u_0} * g_{\gamma^2}\right)(x_1) \\ &= (2\pi\gamma^2)^{p/2} \exp\left\{-\tilde{u}_0(x_1; \gamma^2)\right\} \ .\end{aligned}$$

The Gibbs sampler is covered in Chapter 15.

The prior proximal generator in practice

Use the blurred energy $\tilde{u}_0(x;\sigma^2)$, and expand it to first order about v:

$$\begin{aligned}
\tilde{q}_0(x|v;\sigma^2) &= \frac{1}{Z(v)} \exp\left\{-\tilde{u}_0(x;\sigma^2) - \frac{1}{2\gamma^2}\|x-v\|^2\right\} \\
&\approx \frac{1}{Z'(v)} \exp\left\{(x-v)^t s(v;\sigma^2) - \frac{1}{2\gamma^2}\|x-v\|^2\right\} \\
&= \frac{1}{Z''(v)} \exp\left\{-\frac{1}{2\gamma^2}\left\|x-\left[v+\gamma^2 s(v;\sigma^2)\right]\right\|^2\right\} \ .
\end{aligned}$$

Tweedie's formula and $\beta = \gamma^2/\sigma^2$ then give

$$\tilde{F}_0(v;\beta,\sigma) \approx (1-\beta)\, v + \beta\, \text{Denoise}(v;\sigma) + \sqrt{\beta}\,\sigma W \ ,$$

where $W \sim N(0, I)$.

The approximation is accurate when $\gamma \ll \sigma$, so we choose $\beta < 1$.

The forward proximal generator in practice

Let $\bar{F}_1$ be the ordinary proximal map of the forward model:

$$\bar{F}_1(v;\gamma) = \arg\min_x \left\{ u_1(x) + \frac{1}{2\gamma^2}\|x - v\|^2 \right\} .$$

A first-order expansion of u_1 about $\bar{F}_1(v;\gamma)$ gives, for small γ,

$$q_1(x|v) \approx \frac{1}{Z} \exp\left\{ -\frac{1}{2\gamma^2}\|x - \bar{F}_1(v;\gamma)\|^2 \right\} .$$

The forward proximal generator is the proximal map plus white Gaussian noise:

$$F_1(v;\gamma) \approx \bar{F}_1(v;\gamma) + \gamma W , \qquad W \sim N(0, I) .$$

The GPnP algorithm

GPnP$(\alpha, \beta, \sigma_{max}, \sigma_{min}, N)$

$X \leftarrow \sigma_{max}\,\mathrm{RandN}(0, I) + 1/2$

$a \leftarrow (\sigma_{min}/\sigma_{max})^{1/N}$

For $n = 0$ to $N - 1$ $\{$

$\sigma \leftarrow a^n \sigma_{max}$

$X \leftarrow (1 - \beta)X + \beta\,\mathsf{Denoise}(X; \alpha\sigma) + \sqrt{\beta}\,\sigma\,\mathrm{RandN}(0, I)$

$X \leftarrow \bar{F}_1(X; \sqrt{\beta}\,\sigma) + \sqrt{\beta}\,\sigma\,\mathrm{RandN}(0, I)$

$\}$

Return(X)

- Updates: prior generator, then forward ($\gamma = \sqrt{\beta}\,\sigma$).
- σ decreases geometrically: this is **annealing**.

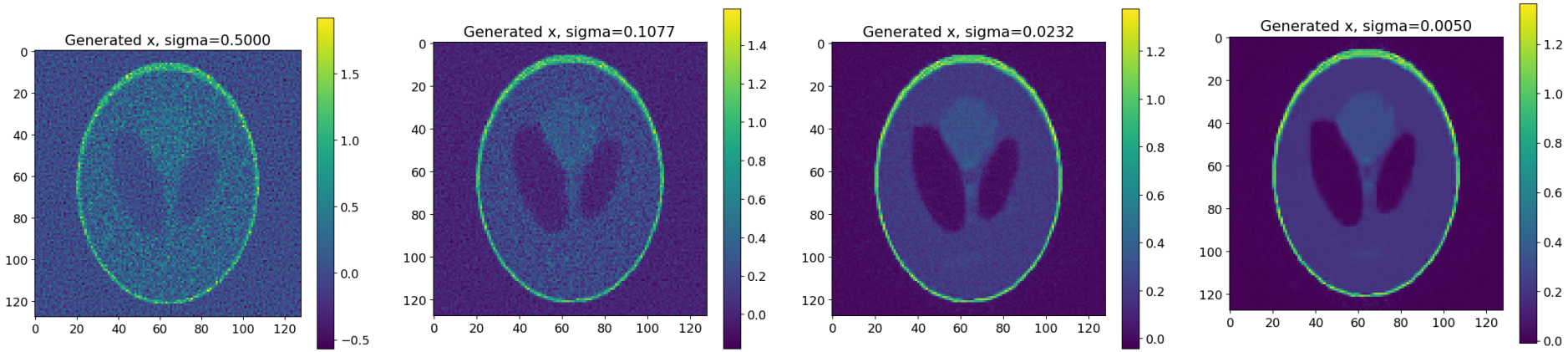
The GPnP parameters

- With the blurred prior $\tilde{u}_0(x;\sigma^2)$, GPnP samples the posterior with the prior blurred by $\sigma^2+\gamma^2$. Small σ at the end makes this blur small.
- Large starting σ lets X move stably through rare regions.

Parameter	Role	Value used
$\beta=\gamma^2/\sigma^2$	step size, $\beta<1$	0.25
$\alpha\in[1,1.5]$	scales the noise level in Denoise$(X;\alpha\sigma)$	1.3
σ_{max}	starting noise level	0.5
σ_{min}	final noise level	0.005
N	number of iterations	100

The parameter α corrects for errors in the denoiser's calibration, and for the decrease in σ on each iteration.

Sparse-view CT: one sample during annealing



$\sigma = 0.500$ $\sigma = 0.107$ $\sigma = 0.023$ $\sigma = 0.005$

- At large σ the prior is near white noise; the image is noisy.
- As σ decreases, the image settles to a less noisy image.
- Each run ends at a different image: that spread is the posterior.

With no data, GPnP is annealed Langevin dynamics

Set $u_1(x) = 0$. Then $F_1(v; \gamma) = v + \gamma W$ exactly, and for small γ,

$$\tilde{F}_0(v) = v + \gamma^2 s(v; \sigma^2) + \gamma W' \ .$$

Composing the two generators gives the **Langevin update**

$$X_n = X_{n-1} + \gamma^2 s(X_{n-1}; \sigma^2) + \sqrt{2}\,\gamma W \ ,$$

or, in terms of the denoiser,

$$X_n = (1 - \beta) X_{n-1} + \beta\,\mathsf{Denoise}(X_{n-1}; \sigma) + \sqrt{2\beta}\,\sigma W \ .$$

With σ decreasing, this is annealed Langevin dynamics. So GPnP with a null forward model is the sampler behind score-based diffusion models.

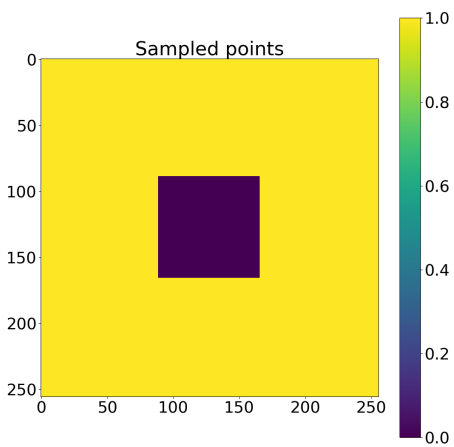
How GPnP differs from diffusion methods

GPnP differs from generative diffusion methods in three ways:

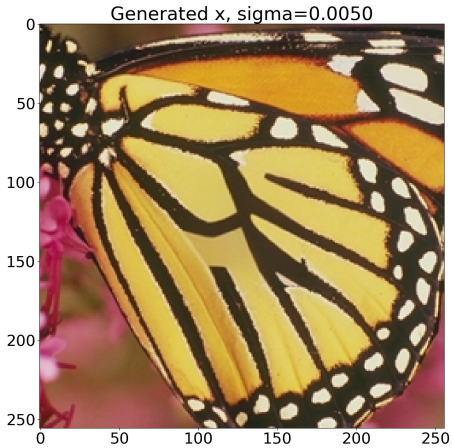
1. GPnP defines its samples by the stationary distribution of a discrete-time Markov chain, not by a stochastic differential equation.
2. Langevin dynamics is used only for a null forward model.
3. GPnP applies proximal generators rather than gradient steps.

- Keeps PnP's modularity: any forward model and MMSE denoiser.
- Other methods also combine a diffusion prior with a physics-based forward model. One example is diffusion posterior sampling (DPS).

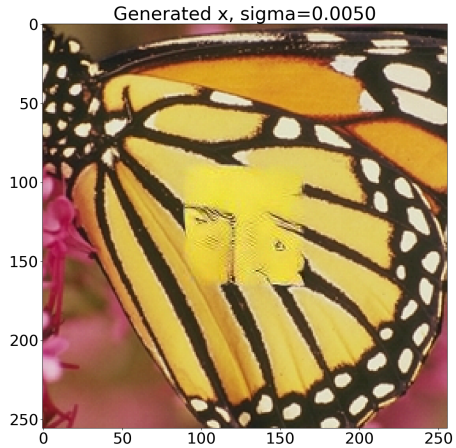
Inpainting with a prior trained on faces

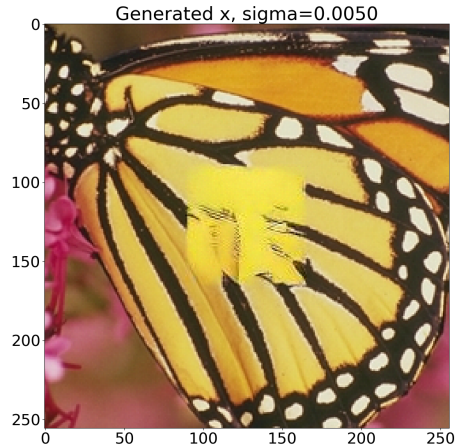


Measured pixels



DRUNet denoiser





Two samples, denoiser trained on face images (CelebA-HQ)

- The data contains no pixels in the center rectangle.
- Face-trained samples fill the rectangle with features not in the truth.

When the data does not constrain a region, the prior determines the samples in that region.

References

- C. A. Bouman and G. T. Buzzard, "Generative Plug and Play: Posterior Sampling for Inverse Problems," Allerton Conference on Communication, Control, and Computing, 2023.
- S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play Priors for Model Based Reconstruction," IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2013.
- G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-Play Unplugged: Optimization-Free Reconstruction Using Consensus Equilibrium," SIAM Journal on Imaging Sciences, 2018.
- U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging: Theory, Algorithms, and Applications," IEEE Signal Processing Magazine, 2023.
- P. Vincent, "A Connection Between Score Matching and Denoising Autoencoders," Neural Computation, 2011.
- Y. Song and S. Ermon, "Generative Modeling by Estimating Gradients of the Data Distribution," Advances in Neural Information Processing Systems (NeurIPS), 2019.
- H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion Posterior Sampling for General Noisy Inverse Problems," International Conference on Learning Representations (ICLR), 2023.

Summary

- Samples $X^{(k)} \sim p_{x|y}(x|y) = \frac{1}{Z}\exp\{-u_1(x) - u_0(x)\}$ show, through their mean and standard deviation, which features the data supports.
- Tweedie's formula $\text{Denoise}(x;\sigma) = x + \sigma^2 s(x;\sigma^2)$ gives the score of the blurred prior $\tilde{p}_{0,\sigma^2} = p_0 * g_{\sigma^2}$.
- The proximal generator draws $F_i(v) \sim q_i(x|v) \propto \exp\{-u_i(x) - \frac{1}{2\gamma^2}\|x - v\|^2\}$, and $X_n = F_1(F_0(X_{n-1}))$ is reversible with stationary distribution $\tilde{p}_{\gamma^2}(x) \propto \exp\{-u_1(x) - \tilde{u}_0(x;\gamma^2)\}$.
- In practice, $\tilde{F}_0(v) \approx (1-\beta)v + \beta\,\text{Denoise}(v;\sigma) + \sqrt{\beta}\,\sigma W$ and $F_1(v) \approx \bar{F}_1(v;\gamma) + \gamma W$ with $\gamma = \sqrt{\beta}\,\sigma$, and σ is annealed from σ_{max} to σ_{min}.
- With $u_1 = 0$, GPnP is annealed Langevin dynamics $X_n = (1-\beta)X_{n-1} + \beta\,\text{Denoise}(X_{n-1};\sigma) + \sqrt{2\beta}\,\sigma W$, and where the data does not constrain the image, the samples come from the prior.

Next lecture: Chapter 12, the EM algorithm.