Ch 2 : Optimization Task

#2.1 Bayesian Estimation Perspective

* Opt. Prob. 형식

$$\hat{x} = \underset{x}{\arg\min} \left(DC(x,y) + \lambda R(x) \right)$$

- 여기서 DC(x,y)는 데이터 일관성항 - 후보 신호 x가 측정 y와 일치할때 작다.
- R(x)는 정규화항, prior 에 따라 x가 있을 가능성이 높을때, 작다
- 정규화 매개변수 $\lambda \geq 0$ 는 데이터 일관성과 정규화항 간의 균형을 조절한다

모델 $y = Ax + e$ 를 사용해서 noise를 고려하자.
Bayesian 관점, likelihood + prior 결합해서
posterior distribution 을 얻고, 그 중 가장 확률높은
x를 선택하는 MAP 로 문제를 풀어보자.
(그런데 알고보니 MAP Estimation이 regularized-optimization 과 수학적으로 동일하네?)

* Bayesian Estimation
- 목표: 측정 y를 기반으로 신호 x를 추정하는 것
- Bayes' Rule 적용해서, MAP (Maximum A Posteriori) Estimator를 유도할 수 있다.

$$\hat{x} = \underset{x}{\arg\max}\ p(x|y) = \underset{x}{\arg\max} \frac{p(y|x)P(x)}{P(y)} = \underset{x}{\arg\max}\ p(y|x)P(x) = \underset{x}{\arg\min} \left[\log p(y|x) - \log p(x) \right]$$

조건부, Bayes 룰 / y는 상관X / 로그를 씌워 계산을 쉽게하자

→ e.g) log-likelyhood term in inverse Problem

log-likelihood 가장 common: Least Square Loss . data가 $y = Ax_*$ 라고 가정했을때 $x_* \sim N(0, \sigma^2 I)$
e는 Gaussian noise, $p(y|x)$ is Gaussian with mean Ax, co-variance matrix $\sigma^2 I$
그럼: $P(y|x) \sim \exp\left(-\frac{1}{2\sigma^2}\|Ax-y\|_2^2\right)$ 이고 → log-likelihood : $-\log(p(y|x)) = \frac{1}{2\sigma^2}\|Ax-y\|_2^2 + C$.

- 측정값 $y = Ax + e$ 일때 데이터 일관성항 ⇒ $DC(x,y) = \frac{1}{2}\|Ax-y\|_2^2$ 이다.
- 여기서 신호가 희소하다고 가정하면 정규화 항은, ℓ_1-Norm 사용 가능 ⇒ $R(x) = \|x\|_1$

따로보기) 밝은 환경의 디카에서 노이즈를 Gaussian 근사하는 것 매우 good.

- digital camera measurement noise
  1) Photon noise (광자 노이즈)
  빛 자체의 노이즈, 물리적 측정한계에서 기인함. Poisson distribution
  2) Sensor-related noise (센서 노이즈)
  회로, OpAmp 등 전자장비에서 발생, 보통 additive Gaussian noise

밝은 환경에서 photon (광자) 수 많음 → Poisson ≈ Gaussian 근사 가능 → noise 전체가 Gaussian이라고 가정해도 OK

BUT)
CT, PET 등 저조도 환경 (dark condition) 에서 photon 수 ⇓ → Poisson 분포 그대로 고려 or 다른 분포가 정확한 순으로
Laplace, impulse, multiplicative etc

왜?
Gaussian noise → log-likelihood는 least-squares loss와 일치 :
$-\log(p(y|x)) \sim \|Ax-y\|_2^2$

* Bayesian Estimation

$$\hat{x} = \underset{x}{\operatorname{argmax}}\ p(x|y) = \underset{x}{\operatorname{argmax}} \frac{p(y|x)\,p(x)}{p(y)} = \underset{x}{\operatorname{argmax}}\ p(y|x)p(x) = \underset{x}{\operatorname{argmin}} \left[\log p(y|x) - \log p(x) \right]$$

조건부, Bayes룰

y는 상관X

로그를 씌워 계산을 쉽게하자

$\hat{x} = \underset{x}{\operatorname{argmax}}\ (\quad\quad) \rightarrow \hat{x} = \underset{x}{\operatorname{argmin}} -\log p(y|x) - \log p(x)$ 형태로 바꾸면 $-\log p(x)$ 가

- 우리가 x에 대해 알고 있는 사전 정보를 나타냄.
- opt관점에서 regularization term

$-\log p(x) \doteq \lambda\sigma^2 R(x)$로 쓰면

$$\hat{x} = \underset{x}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|Ax-y\|_2^2}_{\text{likelihood}} + \underbrace{\lambda}_{\text{trade-off}} \overbrace{R(x)}^{\text{regularizer}}$$

하고싶은말)

1. Bayesian prior는 수학적으로 regularizer로 해석가능
2. MAP estimator는 결국 regularized loss minimization
3. inverse problem이 ill-posed할때 안정적 해를 찾는데 필수조건이다.

그럼 regularizer는 뭘 쓰는데?

#2.2 Common Regularizers

① ℓ_2 Regularizations : $R(x) = \|x\|_2^2$ a.k.a. Tikhonov Regularization, 과도하게 부드러운 이미지로 이어질수있다.

classical — MAP with Gaussian Prior

② ℓ_1 Regularizations : $R(x) = \|x\|_1$ → sparsity-promoting, linked to Laplace prior

③ Total Variation (TV) Regularization :

$R(x) = \sum_i |x_i - x_{i+1}|$

- 이미지의 날카로운 가장자리를 유지하는데 효과적 (can retain sharp edges)
- 가속 자기공명영상? (accelerated magnetic resonance image)에 good!

④ Sparse Regularization :

$R(x) = \|Dx\|_1$

- 특정 기저에서 신호가 희소할때 사용
- 파동기저 이용한 희소정규화(sparsifying basis)도 natural images에 good!

⋮

2.3 Enhancing Deconvolution Stability through Tikhonov Regularization

Idea: Regularization은 단순히 "좋은 해"를 얻기 위한 수단 뿐만 아니라 Stability를 확보할때 필수적.
inverse Problem이 noise에 매우 민감함 (ill-posed), 그리고 Regularized problem은 다음 성질을 가짐
① Solution exists ② Solution unique ③ Solution stable
(small noise → small change in solution)

trade-off
- Regularization 너무 세면 → bias 발생
- 너무 약하면 → variance 커짐

* e.g) 1-dim Deconvolution Problem

- 목표: noise가 포함된 $y = Ax + e$ 로 부터 신호 x 복원하기.
- consider the least-square estimate $\hat{x}_{LS}(y) = \underset{x}{\arg\min}\|Ax-y\|_2^2$
- 일반적으로 $\hat{x}_{LS}(y) = A^{-1}y = x + V\Sigma^{-1}U^Te$ 로 쓸 수 있다, $A = U\Sigma V^T$는 SVD (Singular Value Decomposition)
- noise $e=0$: LSE는 신호 x를 완벽 복구 가능
- noise $e \neq 0$: A가 poorly conditioned하므로 가장 작은 특이값에 대한 가장 큰 값의 비율이 매우 크다
(the fraction of largest over smallest singular value is very large)

error)
$$E\left[\|\hat{x}_{LS}(y) - x\|_2^2\right] = E\left[\|V\Sigma^{-1}U^Te\|_2^2\right] = E\left[\|\Sigma^{-1}U^Te\|_2^2\right]$$
V is unitary

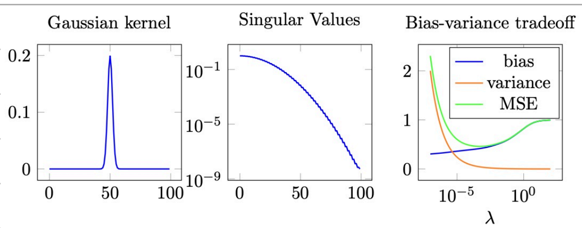
- V orthogonal matrix ⇒ Norm 같음
$\|Vz\|_2 = \|z\|_2$

* Regularized Least-Square Estimate

- 정규화된 estimator :

$$\hat{x}_\lambda(y) = \underset{x}{\arg\min}\left(\|Ax-y\|_2^2 + \lambda\|x\|_2^2\right) \xrightarrow[\text{form}]{\text{closed-}} \hat{x}_\lambda(y) = V\,\mathrm{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}, \cdots, \frac{\sigma_d^2}{\sigma_d^2+\lambda}\right)V^Tx + V\,\mathrm{diag}\left(\frac{\sigma_1}{\sigma_1+\lambda}, \cdots, \frac{\sigma_d}{\sigma_d+\lambda}\right)U^Te$$

- Gaussian kernel의 Deconvolution Problem에 대한 bias-trade off 그래프



→ (Right graph) The expected mean-squared reconstruction

$$E\left[\|\hat{x}_\lambda(y) - x\|_2^2\right] = \underbrace{\sum_{i=1}^{n}\left(1-\frac{\sigma_i^2}{\sigma_i^2+\lambda}\right)^2 (v_i^Tx)^2}_{\text{bias}} + \underbrace{\sigma^2\sum_{i=1}^{n}\left(\frac{\sigma_i^2}{\sigma_i^2+\lambda}\right)^2}_{\text{variance}}$$

★ bias-variance trade off를 통해 최적의 파라미터 λ를 선택해 더 나은 신호 estimate을 할 수 있음.

L2 vs L1 직관)

$R(x) = \|x\|_2^2 = \sum x_i^2$	$R(x) = \|x\|_1 = \sum \lvert x_i\rvert$
$Loss(x) = \min \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|x\|_2^2$	$Loss(x) = \min \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|x\|_1$
모든 변수 제곱합	
크기가 큰 값 억제	희소성(sparsity) 유도
모든 값을 작게 만든다 → 해가 dense (부드럽게 다 줄여)	어떤 값은 아예 0으로 만든다 → 해가 sparse